A Comparison of Two Heavy Traffic Regimes:

I: Multiserver Queues

II: Multiserver Queueing Networks

Martin I. Reiman

Bell Labs, Murray Hill, NJ, USA

ABSTRACT

The traffic intensity, ρ, of a single station queueing system provides a simple measure of the ability of the system to handle the arriving work. Roughly speaking, if $\rho < 1$ then the arrival rate is smaller than the service rate and the system is stable, while if $\rho > 1$, then the system is unstable. Heavy traffic limit theorems consider a sequence of systems with $\rho \to 1$, and yield both important insights and useful approximations.

In this talk we present two heavy traffic regimes for queueing systems. We begin with the *M/M/N* queue, essentially the simplest *N* server queueing system.
In the classical heavy traffic regime for this system *N* is held fixed and $\rho \to 1$. With proper normalization the queue length process converges to a one-dimensional reflected Brownian motion (as shown in the more general *GI/G/N* setting by Iglehart and Whitt [3]). The second regime, introduced by Halfin and Whitt [1], has $N \to \infty$ as $\rho \to 1$, with $(1- \rho_N)\sqrt{N} \to \beta$ and $-\infty < \beta < \infty$. With the appropriate centering and normalization the queue length here converges to a one-dimensional diffusion obtained by 'pieceing together' a Brownian motion and an Ornstein-Uhlenbeck process.

As noted above, the heavy traffic limit theorem of [3] actually covers the more general *GI/G/N* queueing system. A heavy traffic limit theorem for the *GI/G/N* queue in the 'Halfin-Whitt regime' has not yet been proved: non-exponential service times pose a fundamental difficulty in this regime. A proof for the *GI/PH/N* queue, in which service times have a 'phase-type' distribution was provided in [6]. Here the limit process is a multidimensional diffusion process.

Queueing networks consist of several stations, each with its own servers, with customers moving between the stations of the network. The Jackson network (introduced in [4]) consists of *K M/M/N* stations. The traffic intensity for the network, which is a *K* dimensional vector, again can be used to demarcate stable from unstable systems. A limit theorem in the 'classical' heavy traffic regime (applying to 'generalized' Jackson networks) was proved in [7]. The limit process (defined in [2]) is a reflected Brownian motion in the *K* dimensional non-negative orthant. A heavy traffic limit theorem in the Halfin-Whitt regime was proved in [5]. The limit process here is a *K* dimensional diffusion without boundaries.

The limit theorem of [5] actually covers a more general setting, called Markovian service networks in [5]. The key restriction is that service times must be exponential. With this restriction satisfied a variety of networks, including some with multiple classes and priorities are allowed. For multiple class networks it is known that in some instances $\rho_i < 1$ (ρ_i is the traffic intensity of station i) for all stations is not a sufficient condition for stability. An early example of this is provided by Rybko and Stolyar [8]. It is thus not surprising that classical heavy traffic limit theorems for general multiclass networks have not been obtained. What is perhaps surprising is that the limit theorem of [5] applies to the Rybko-Stolyar network.

References

[1] Halfin, S. and W. Whitt: *Heavy-Traffic Limits for Queues with Many Exponential Servers,* Operations Research **29** (1981) 567 - 588**.**

[2] Harrison, J. M. and M.I. Reiman, *Reflected Brownian Motion on an Orthant,* Annals of Probabability **9** (1981) 302 - 308.

[3] Iglehart, D. L. and W. Whitt: *Multiple Channel Queues in Heavy Traffic, II: Sequences, Networks, and Batches,* Advances in Applied Probability **2** (1970) 355 - 364.

[4] Jackson, J. R.: *Networks of Waiting Lines,* Operations Research **5** (1957) 518 – 521.

[5] Mandelbaum, A., W.A. Massey and M.I. Reiman: *Strong Approximations for Markovian Service Networks,* Queueing Systems: Theory and Applications (QUESTA) **30** 149 - 201.

[6] Puhalskii, A. A. and M.I. Reiman, M. I.: *The Multiclass GI/PH/N Queue in the Halfin-Whitt Regime,* Advances in. Applied Probability **32** (2000) 564 - 595.

[7] Reiman, M. I.: *Open Queueing Networks in Heavy Traffic,* Mathematics of Operations Research 9 (1984) 441 - 458.

[8] Rybko, A. N. and A.L. Stolyar: *Ergodicity of Stochastic Processes Describing the Operation of an Open Queueing Network,* Problem of Information Transmission **27** (1992).