Heavy traffic steady state approximations in stochastic networks with Lévy inputs

Jean-Paul Haddad
Department of Electrical and Computer Engineering
University of Waterloo
200 University West
Waterloo, Ontario, Canada
N2L 3G1
jhaddad@uwaterloo.ca

Ravi R. Mazumdar
Department of Electrical and Computer Engineering
University of Waterloo
200 University West
Waterloo, Ontario, Canada
N2L 3G1
mazum@ece.uwaterloo.ca

ABSTRACT

It has recently been shown [3, 5] that in the heavy traffic limit, the stationary distributions of the scaled queue length process of Generalized Jackson Networks converges to the stationary distribution of its corresponding Reflected Brownian Motion limit. In this paper we show that such an "interchange of limits" is valid for the workload process of Stochastic Fluid Networks with Lévy inputs. Our technique is of independent interest because we do not require the use of any Lyapunov techniques, a method that was used in the previous two papers.

Keywords

Diffusion approximations, Weak convergence, Stationary distribution, Reflected brownian motion

1. INTRODUCTION

It is often the case in the performance analysis of stochastic networks that the calculation of the stationary distribution (if it exists) of the workload process is of great importance. Unfortunately, besides the case where the system can be modeled as a Jackson or BCMP network, not many closed form solutions exist.

On the other hand heavy traffic analysis often leads to Reflected Brownian Motion (more precisely Semi-martingale Reflected Brownian Motion (SRBM)) for which under suitable hypothesis the stationary distributions can be explicitly computed. Thus diffusion approximations to the workload process, especially Reflected Brownian Motion, have been the focus of much recent research.

The issue of the stability of networks has been a preoccupation of the stochastic network and queueing community for quite some time and except for successes in the Markovian setting or for i.i.d. inputs, there are not many results available for the general case. The recent monograph of Bramson [2] aptly summarizes the state of the art but essentially restricts itself to the i.i.d or generalized Jackson case.

VALUETOOLS 2009, October 20-22, Pisa, Italy

DOI 10.4108/ICST.VALUETOOLS2009.7463

Yet it can be seen that there are a plethora of papers that take for granted that the intuitive condition that the average load at a queue being less than the server capacity under any work conserving discipline implies the existence of a stationary distribution. This remains an open question in general.

This raises the related question. Does the existence of a stationary or invariant distribution for a diffusion limit mean that the original network also is stable? In Chen & Yao [4, 14] they point out the fact that the stationary distribution of Reflected Brownian Motion is often used as an approximation to the scaled stationary distribution of the workload process for Generalized Jackson networks.

In this paper we justify this interchange of limits for Stochastic Fluid Networks with Lévy inputs. This conjecture was rigorously justified by Gamarnik & Zeevi [5] for Generalized Jackson networks using Lyapunov function techniques assuming that all moments for the input sequences exist and later extended by Budhiraja & Lee [3] who showed that only second order assumptions are sufficient also using Lyapunov techniques. In this paper we show the corresponding results for networks with Lévy inputs using different arguments, but still only requiring second order assumptions.

Stochastic fluid networks are natural models for studying systems where the inputs can not be enumerated and good approximations for queueing systems such as high speed communication networks. In this paper, we will consider an open single class, single server stochastic fluid network with independent Lévy inputs, which is a generalization of an open network of $M/G/\cdot$ input queues. This model has been extensively studied in a series of papers by Kella [9], Kella & Whitt [12, 11] and in [13].

Diffusion, or heavy traffic approximations have been the focus of vigorous research for a long time. Reflected Brownian motion (RBM) was introduced by Harrison & Reiman [6], Harrison & Williams [7]. The convergence of networks in the heavy traffic limit to SRBM is by now well known. See also the survey in [16]. The monograph of Whitt is a comprehensive reference [15].

2. ASSUMPTIONS AND NOTATION

Fix an integer $N \geq 0$ and let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a stochastic basis. All processes will be assumed to be $(\mathcal{F}_t)$-adapted and in $D[0,\infty)$ (the space of N-dimensional, real-valued, càdlàg processes). The space $D[0,\infty)$ will be assumed to be endowed with the Skorokhod J_1 topology, unless explicitly stated otherwise.

Vectors and matrices are assumed to be real-valued, with vectors being column vectors. The expression $f_n \sim g_n$ means $\lim_{n\to\infty} \frac{f_n}{g_n} = K < \infty$ for some constant K, and $\mathcal{L}(X)$ will denote the law of the random quantity X.

3. THE STOCHASTIC FLUID NETWORK

The Stochastic fluid networks in this paper are assumed to be open networks, that possess N single server queues that are served by constant rate servers. Stochastic fluid networks are characterized by the 4-tuple $\{J, r, P, W(0)\}$ where $\{J(t); t \geq 0\}$ is the cumulative input process, $r \in \mathbb{R}_+^N$ is the server rate, $P \in \mathbb{R}_+^N \times \mathbb{R}_+^N$ the routing matrix, and $W(0)$ is the initial workload. It is assumed that the queues are work conserving.

The routing matrix P is a substochastic matrix such that $P^n \to 0$ as $n \to \infty$. Denote $Q = (I - P')^{-1} = I + P' + P'^2 + \ldots$, P' is the transpose of P. r_i represents the rate at which server i processes fluid (if there is any).

The processes $\{J_i(t); t \geq 0\}$, $i \in 1 \ldots N$ are independent subordinators (non-decreasing càdlàg Lévy processes) with $J(0) = 0$, with finite first and second moments (on bounded time intervals). The process $J_i(t)$ represents the cumulative input to server i at time t. We can assume wlog that J is a pure jump process, since any drift term can be subtracted away from the service rate. Also, we assume $E[W(0)] < \infty$ and that $W(0)$ is independent of $\{J(t); t \geq 0\}$.

Let $\{X(t); t \geq 0\}$ be the net input, ie. $X(t) = W(0) + J(t) - (I - P')rt$. The Skorokhod Oblique Reflection Problem states that given the process X(t) and the matrix I-P', there exists a.s. unique processes $\{W(t); t \geq 0\}$ (known as the reflected process) and $\{Z(t); t \geq 0\}$ (known as the regulator) such that:

1. $W(t) = X(t) + (I - P')Z(t) \geq 0$
2. $Z(0) = 0$ and $dZ(t) \geq 0$
3. $W_i(t)dZ_i(t) = 0$

Furthermore there exists a unique, continuous pair of functions $\Phi, \Psi : D[0,\infty) \to D[0,\infty)$ such that $\Phi(X(\cdot)) = W(\cdot)$ and $\Psi(X(\cdot)) = Z(\cdot)$. See Chapter 14 of Whitt [15] or Chapter 7 of Chen & Yao [14] for further details. The workload process will be represented by $\{W(t); t \geq 0\}$.

Let $\lambda = E[J(1)]$, where λ_i represents the average rate of work arriving at server i per unit of time, $\sigma^2 = Var(J(1))$ and let $\Gamma = diag(\sigma^2)$ denote the covariance matrix of J.

We will assume the stability conditions $r > Q\lambda$ and $E[W(0)] < \infty$. From Kella [10], these conditions are necessary and sufficient for the existence of a unique stationary distribution for $\{W(t); t \geq 0\}$, with $W(\infty)$ being a random vector with the stationary distribution as its law. Note that if $W(0) \stackrel{\mathcal{D}}{=} W(\infty)$, then $\{W(t); t \geq 0\}$ is in steady state.

4. HEAVY TRAFFIC APPROXIMATION

The main assumption to be made is that for the sequence of stochastic fluid networks $\{J_n, r_n, P, W_n(0)\}$ (which satisfy the previous assumptions $\forall n$),

$$\sqrt{n}(\lambda_n - (I - P')r_n) \to \eta$$

.
Note that we have $\eta < 0$, since $r_n > Q\lambda_n$ for all n.

We also assume that $\frac{W_n(0)}{\sqrt{n}}$ converges in distribution to some random vector W_0. Additionally, we require that the sequence $\{J_n^2(1)\}$ is uniformly integrable.

The heavy traffic limit will be shown to be Reflected Brownian Motion, denoted by $RBM_X(b, \Gamma, R)$ where X is the initial distribution, b is the drift, Γ is the covariance matrix, and R is the reflection matrix.

We require the following result from Harrison & Williams [7]: The $RBM_X(b, \Gamma, R)$ possesses a unique stationary distribution if $R^{-1}b < 0$.

5. THE MAIN RESULT

The outline of the approach to show the result is straightforward. We will show that, independent of (reasonable) initial conditions, the heavy traffic limit of $\frac{W_n(nt)}{\sqrt{n}}$ is Reflected Brownian Motion. Then we will show that the scaled sequence of stationary distributions is tight. Finally we will prove the interchange result.

5.1 Convergence to Reflected Brownian Motion

LEMMA 5.1. *Let*

$$\bar{J}_n(t) = \frac{J_n(nt) - \lambda_n nt}{\sqrt{n}}$$

Then

$$\overline{J}_n(\cdot) \Rightarrow BM(0, \Gamma)$$

Where $BM(0, \Gamma)$ denotes a 0 drift Brownian motion with covariance matrix Γ.

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7463
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7463

PROOF. By the independent and stationary increments property of $J_n(t)$,

$$J_n(nt) = \sum_{i=1}^{n} \widetilde{J_n^i}(t)$$

where $\widetilde{J_n^i}(t)$ are independent copies of $J_n(t)$.

Thus, by the (triangular array) CLT,

$$\overline{J}_n(1) \Rightarrow N(0,\Gamma)$$

By VII Corollary 3.6 of Jacod & Shiryaev [8],

$$\overline{J}_n(\cdot) \Rightarrow BM(0,\Gamma)$$

□

THEOREM 5.1. *Let*

$$\overline{W}_n(\cdot) = \frac{W_n(n\cdot)}{\sqrt{n}}$$

Then

$$\overline{W}_n(\cdot) \Rightarrow RBM_{W_0}(\eta,\Gamma,I-P')$$

PROOF. From Theorem 5.1, we know

$$\overline{J}_n(\cdot) \Rightarrow BM(0,\Gamma)$$

From the assumptions,

$$\frac{W_n(0)}{\sqrt{n}} \Rightarrow W_0$$

and

$$\sqrt{n}(\lambda_n - (I-P')r_n) \to \eta$$

Since $W_n(0)$ and $J_n(t)$ are independent, and η is deterministic, then they converge jointly.

Let $\overline{X}_n(t) = \frac{X_n(nt)}{\sqrt{n}}$

$$\begin{aligned}\overline{X}_n(t) &= \frac{W_n(0)}{\sqrt{n}} + \frac{J(nt)}{\sqrt{n}} - \frac{(I-P')r_n nt}{\sqrt{n}}\\ &= \frac{W_n(0)}{\sqrt{n}} + \frac{J(nt)-\lambda_n nt}{\sqrt{n}} + \sqrt{n}(\lambda_n-(I-P')r_n)t\\ &= \frac{W_n(0)}{\sqrt{n}} + \overline{J}_n(t) + \sqrt{n}(\lambda_n-(I-P')r_n)t\end{aligned}$$

Therefore $\overline{X}_n(\cdot) \Rightarrow W_0 + BM(\eta,\Gamma)$

$$\begin{aligned}\overline{W}_n(\cdot) &= \Phi(\overline{X}_n(\cdot))\\ &\Rightarrow RBM_{W_0}(\eta,\Gamma,I-P')\end{aligned}$$

By the continuous mapping theorem, since the mapping Φ is continuous by assumption.

As well,

$$\begin{aligned}\overline{W}_n(\cdot) &= \frac{\Phi(X_n(n\cdot))}{\sqrt{n}}\\ &= \Phi(\frac{X_n(n\cdot)}{\sqrt{n}})\end{aligned}$$

by uniqueness of the mapping Φ.

Therefore,

$$\overline{W}_n(\cdot) \Rightarrow RBM_{W_0}(\eta,\Gamma,I-P')$$

□

5.2 Tightness

LEMMA 5.2. *There exists $\tilde{\lambda}_n$ s.t. $\lambda_n < \tilde{\lambda}_n$, $Q\tilde{\lambda}_n < r_n$ and $\tilde{\lambda}_n - \lambda_n \sim \frac{1}{\sqrt{n}}$*

PROOF. Let $\epsilon_n = \frac{\min_{i=1\ldots N}(-Q(\lambda_n-(I-P')r_n))_i}{2N\|Q\|_{\max}} e$ and set $\tilde{\lambda}_n = \lambda_n + \epsilon_n$.

Where $\|\cdot\|_{\max}$ is the max norm and e is a column vector of ones.

By assumption $\epsilon_n > 0$, therefore $\lambda_n < \tilde{\lambda}_n$.

$$\begin{aligned}Q\tilde{\lambda}_n &= Q(\lambda_n + \frac{\min_{i=1\ldots N}(-Q(\lambda_n-(I-P')r_n))_i}{2N\|Q\|_{\max}}e)\\ &= Q\lambda_n + Q\frac{\min_{i=1\ldots N}(-Q(\lambda_n-(I-P')r_n))_i}{2N\|Q\|_{\max}}e\\ &\le Q\lambda_n + \frac{\min_{i=1\ldots N}(-Q(\lambda_n-(I-P')r_n))_i}{2}e\\ &\le Q\lambda_n - \frac{1}{2}Q(\lambda_n-(I-P')r_n)\\ &= \frac{1}{2}Q\lambda_n + \frac{1}{2}r_n\\ &< \frac{1}{2}r_n + \frac{1}{2}r_n\\ &= r_n\end{aligned}$$

Finally we show the last assertion.

$$\begin{aligned}\frac{\tilde{\lambda}_n-\lambda_n}{\frac{1}{\sqrt{n}}} &= \sqrt{n}\frac{\min_{i=1\ldots N}(-Q(\lambda_n-(I-P')r_n))_i}{2N\|Q\|_{\max}}e\\ &= \frac{\min_{i=1\ldots N}(-Q\sqrt{n}(\lambda_n-(I-P')r_n))_i}{2N\|Q\|_{\max}}e\\ &\to \frac{\min_{i=1\ldots N}(-Q\eta)_i}{2N\|Q\|_{\max}}e\end{aligned}$$

Note that $\frac{\min_{i=1\ldots N}(-Q\eta)_i}{2N\|Q\|_{\max}} > 0$ by assumption. □

LEMMA 5.3. $\sum_{j=1}^{N} E[W_{n,j}(\infty)] \le \sum_{j=1}^{N} \frac{\sigma_{n,j}^2}{2(\tilde{\lambda}_n-\lambda_n)_j}$ *where $\tilde{\lambda}_n$ satisfies the conditions of Lemma 5.2.*

PROOF. Fix n.

Consider the reflected process

$$\tilde{W}_n(\cdot) = W_n(0) + J_n(\cdot) - \tilde{\lambda}_n \cdot + \tilde{Z}_n(\cdot)$$

Since $\lambda_n < \tilde{\lambda}_n$, there a exists a unique stationary distribution for the process $\{\tilde{W}_n(t); t \ge 0\}$.

Let $\tilde{W}_n(\infty)$ be a random vector with that stationary distribution.

From [12], for each $i = 1\ldots N$,

$$\tilde{W}_{n,i}(\infty) = \frac{\sigma_{n,i}^2}{2(\tilde{\lambda}_n-\lambda_n)_i}$$

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7463
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7463

By Lemma 3.1 of Kella [10],

$$\sum_{j=1}^{N} W_{n,j}(t) \leq \sum_{j=1}^{N} \tilde{W}_{n,j}(t)$$

This implies that

$$E[\sum_{j=1}^{N} W_{n,j}(\infty)] \leq E[\sum_{j=1}^{N} \tilde{W}_{n,j}(\infty)]$$

The result follows. □

LEMMA 5.4. *There exists a finite constant A s.t.* $\frac{E[W_{n,k}(\infty)]}{\sqrt{n}} \leq A \;\; \forall k = 1 \dots N$

PROOF. Let $q_{max} = \|Q\|_{\max}$, $\sigma_n^2 = \max_j(\sigma_{n,j}^2)$

$$\sum_{j=1}^{N} E[W_{n,j}(\infty)] \geq E[W_{n,k}(\infty)] \;\; \forall k = 1 \dots N$$

by the non-negativity of the summands and

$$\sum_{j=1}^{N} \frac{\sigma_{n,j}^2}{2(\tilde{\lambda}_n - \lambda_n)_j} \leq \sum_{j=1}^{N} \frac{\sigma_n^2}{2(\tilde{\lambda}_n - \lambda_n)_j}$$

By Lemma 5.2, $\forall j$, $(\tilde{\lambda}_n - \lambda_n)_j \sim \frac{1}{\sqrt{n}}$

$$\begin{aligned} & \frac{\sum_{j=1}^{N} \frac{\sigma_n^2}{2(\tilde{\lambda}_n - \lambda_n)_j}}{\sqrt{n}} \\ = & \sum_{j=1}^{N} \frac{\sigma_n^2}{2\sqrt{n}(\tilde{\lambda}_n - \lambda_n)_j} \\ \to & \sum_{j=1}^{N} N q_{\max} \frac{\sigma_n^2}{\min_{i=1\dots N}(-Q\eta)_i} \\ = & N^2 q_{\max} \frac{\sigma_n^2}{\min_{i=1\dots N}(-Q\eta)_i} \end{aligned}$$

Let

$$A_n = \frac{\sum_{j=1}^{N} \frac{\sigma_n^2}{2(\tilde{\lambda}_n - \lambda_n)_j}}{\sqrt{n}}$$

and

$$A = \sup_n A_n$$

Since $\forall n \; A_n < \infty$ and

$$\lim_{n\to\infty} A_n = N^2 q_{\max} \frac{\sigma_n^2}{\min_{i=1\dots N}(-Q\eta)_i} < \infty$$

So $A < \infty$ and

$$\frac{E[W_{n,k}(\infty)]}{\sqrt{n}} \leq A \;\; \forall k = 1 \dots N$$

□

THEOREM 5.2. *The sequence of stationary distributions* $\{\mathcal{L}(\frac{W_n(\infty)}{\sqrt{n}})\}$ *is tight.*

PROOF. For any $K > 0$,

$$\begin{aligned} P(\frac{W_n(\infty)}{\sqrt{n}} > K) &= P(W_n(\infty) > K\sqrt{n}) \\ & \quad \text{By the union bound} \\ &\leq \sum_{j=1}^{N} P(W_{n,j}(\infty) > K\sqrt{n}) \\ & \quad \text{By the Markov inequality} \\ &\leq \sum_{j=1}^{N} \frac{E[W_{n,j}(\infty)]}{K\sqrt{n}} \\ &\leq \sum_{j=1}^{N} \frac{A}{K} \text{ by Lemma 5.4} \\ &= \frac{NA}{K} \end{aligned}$$

Tightness follows from the inequality above. □

Before proving the main result of the paper, we require the following facts about tight sequences of measures. Proofs can be found in Billingsley [1] (pg. 59, Theorem 5.1 and its Corollary)
Let $\{\pi_n\}$ be a sequence of tight measures. Then:

- Every subsequence contains a weakly convergent subsequence
- If each convergent subsequence converges to the same measure π^*, then $\{\pi_n\} \Rightarrow \pi^*$

We now complete the proof of the main result on the interchange.

THEOREM 5.3. *The sequence* $\{\mathcal{L}(\frac{W_n(\infty)}{\sqrt{n}})\}$ *converges weakly to* π_{RBM}*, where* π_{RBM} *is the unique stationary distribution of* $RBM_{W_0}(\eta, \Gamma, I - P')$

PROOF. Let $\{\mathcal{L}(\frac{W_{n_k}(\infty)}{\sqrt{n_k}})\}$ be a convergent subsequence with $\mathcal{L}(\frac{W_{n_k}(\infty)}{\sqrt{n_k}}) \Rightarrow \pi$, and let $W_{n_k}(0) \stackrel{\mathcal{D}}{=} W_{n_k}(\infty)$. To simplify the notation, denote $\frac{W_{n_k}(n_k \cdot)}{\sqrt{n_k}}$ by $\overline{W}_{n_k}(\cdot)$.

By Theorem 5.1, $\overline{W}_{n_k}(\cdot) \Rightarrow RBM_{W_0}(\eta, \Gamma, I - P')$, with $\mathcal{L}(W_0) = \pi$. Moreover, since for any fixed $t \geq 0$, $\overline{W}_{n_k}(t) \stackrel{\mathcal{D}}{=} \overline{W}_{n_k}(0)$, then $RBM_{W_0}(\eta, \Gamma, I - P')(t) \stackrel{\mathcal{D}}{=} W_0$. This implies that π is a stationary distribution of the RBM. But the stationary distribution of RBM is unique, so therefore $\pi = \pi_{RBM}$.

Since this was true for any arbitrary convergent subsequence, $\mathcal{L}(\frac{W_n(\infty)}{\sqrt{n}}) \Rightarrow \pi_{RBM}$. □

6. ACKNOWLEDGEMENTS

This research was supported in part by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

7. REFERENCES

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7463
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7463

[1] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999.

[2] M. Bramson. *Stability of queueing networks*, volume 1950 of *Lecture Notes in Mathematics*. Springer, Berlin, 2008.

[3] A. Budhiraja and C. Lee. Stationary distribution convergence for generalized jackson networks in heavy traffic. *Math. of Operations research*, 34(1):45–56, 2009.

[4] H. Chen and D.D. Yao. *Fundamentals of queueing networks*, volume 46 of *Applications of Mathematics*. Springer-Verlag, New York, 2001.

[5] D. Gamarnik and A. Zeevi. Validity of heavy traffic steady-state approximation in generalized Jackson networks. *Ann. Appl. Probab.*, 16(1):56–90, 2006.

[6] J.M. Harrison and M. I. Reiman. Reflected brownian motion in an orthant. *Ann. Probab.*, 9:302–308, 1981.

[7] J.M. Harrison and R. J. Williams. Brownian models of open queueing metworks with homogeneous customer populations. *Stochastics*, 22:77–115, 1987.

[8] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1987.

[9] O. Kella. Parallel and tandem fluid networks with dependent lévy inputs. *Ann. Appl. Probab.*, 3(3):682–695, 1993.

[10] O. Kella. Stability and nonproduct form of stochastic fluid networks with Lévy inputs. *Ann. Appl. Probab.*, 6(1):186–199, 1996.

[11] O. Kella and W. Whitt. Diffusion approximations for queues with server vacations. *Adv. in Appl. Probab.*, 22:706–729, 1990.

[12] O. Kella and W. Whitt. Useful martingales for stochastic storage processes with lévy input. *J. Appl. Probab.*, 29(2):396–403, 1992.

[13] T. Konstantopoulos, G. Last, and S.-J. Lin. On a class of Lévy stochastic networks. *Queueing Syst.*, 46:409–437, 2004.

[14] X. Shen, H. Chen, J. G. Dai, and W. Dai. The finite element method for computing the stationary distribution of an SRBM in a hypercube with applications to finite buffer queueing networks. *Queueing Systems Theory Appl.*, 42(1):33–62, 2002.

[15] W. Whitt. *Stochastic-process limits*. Springer Series in Operations Research. Springer-Verlag, New York, 2002.

[16] R. J. Williams. Semimartingale reflecting Brownian motions in the orthant. In *Stochastic networks*, volume 71 of *IMA Vol. Math. Appl.*, pages 125–137. Springer, New York, 1995.

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7463
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7463