Diffusion Parameters of Flows in Stable Queueing Networks

Yoni Nazarathy
The University of Queensland

Werner Scheinhardt
University of Twente

A Network of Queues

\[
\begin{pmatrix}
0 & 0 & 1/2 & 1/2 & 0 & 0 \\
1/2 & 0 & 0 & 1/2 & 0 & 0 \\
1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/2 & 1/2 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
\]

\[
\begin{pmatrix}
8.25 \\
8.25 \\
5 \\
8.25 \\
5 \\
5 \\
\end{pmatrix}
\]

\[
\begin{pmatrix}
1 \\
4 \\
0 \\
0 \\
0 \\
0 \\
\end{pmatrix}
\]

\[
\begin{pmatrix}
2 \\
2 \\
0 \\
0 \\
0 \\
0 \\
\end{pmatrix}
\]
Network Equations

\[ Q_k(t) = Q_k(0) + A_k(t) + \sum_{i=1}^{K} D_{i,k}(t) - \sum_{j=0}^{K} D_{k,j}(t) \]

\[ D_{i,j}(t) = \Phi_{i,j} \left( S_i(T_i(t)) \right), \quad i = 1, \ldots, K, \quad j = 0, \ldots, K \]

\[ \sum_{j=0}^{K} \Phi_{i,j}(\ell) = \ell, \quad i = 1, \ldots, K \]

\[ D = \begin{bmatrix} D_{1,1}, \ldots, D_{1,K}, D_{2,1}, \ldots, D_{2,K}, \ldots, D_{K,1}, \ldots, D_{K,K} \end{bmatrix}' \]
Asymptotic Rates

\[ \nu_{i \to j} := \lim_{t \to \infty} \frac{E[D_{i,j}(t)]}{t} \]

\[ \sigma_{i \to j}^2 := \lim_{t \to \infty} \frac{\text{Var}(D_{i,j}(t))}{t}, \quad \sigma_{i_1 \to j_1, i_2 \to j_2} := \lim_{t \to \infty} \frac{\text{Cov}(D_{i_1,j_1}(t), D_{i_2,j_2}(t))}{t} \]

Calculating \( \nu_{i \to j} \) is easy

**Our contribution:** Closed formulas for the (co)variance terms with diffusion limits
Fluid and Diffusion Scaling and Limits

For \( n = 1, 2, \ldots \) and a function \( U(t) \), denote \( \bar{U}^n(t) = \frac{U(nt)}{n} \). We say that a fluid limit of \( U \) exists if \( \lim_{n \to \infty} \bar{U}^n(t) = \bar{U}(t) \) exists uniformly on compact sets (u.o.c) almost surely. Further, when the limit \( \bar{U}(t) \) exists, denote,

\[
\hat{U}^n(t) = \frac{U(nt) - \bar{U}(nt)}{\sqrt{n}}, \quad n = 1, 2, \ldots.
\]

In cases where the above sequence converges weakly on Skorohod \( J_1 \) topology to a limiting process, \( \hat{U}(t) \), we denote,

\[
\hat{U}^n \Rightarrow \hat{U}
\]

For discrete time processes replace \( U(nt) \) by \( U([nt]) \)
Assumptions on Primitives

\( A_k(\cdot), S_k(\cdot) \) and \( \Phi_k, \cdot(\cdot) \) independent

**FSLLN:**
\[
\bar{A}_i(t) = \alpha_i t, \quad \bar{S}_i(t) = \mu_i t, \quad \bar{\Phi}_{i,j}(\ell) = p_{i,j}\ell,
\]
with \( \alpha_i > 0, \mu_i > 0, p_{i,j} \geq 0 \), and \( p_{i,0} = (1 - \sum_{j=1}^{N} p_{i,j}) \geq 0 \) and \( \text{sp}(P) < 1 \) so \( \nu_{i \rightarrow j} := (I - P')^{-1} \alpha p_{i,j} \)

**FCLT:**
\( \hat{A}_i(t) \) are BM with coefficients \( \nu_i \geq 0 \)
\[
\hat{\Phi}_k,(t) = \left[ \hat{\Phi}_{k,1}(t), \ldots, \hat{\Phi}_{k,K}(t) \right]', \quad k = 1, \ldots, K,
\]
are BM with cov matrices \( \Gamma_k \), having entries \( p_{k,i}(\delta_{i,j} - p_{k,j}) \)

**UI:**
\[
\left\{ \frac{(A_i(t) - \alpha_i t)^2}{t}, \ t \geq t_0 \right\} \text{ is UI}
\]
Stable Scheduling Policy Assumptions

Different policies imply different restrictions on $T(t)$ (single-class, multi-class, preemptive, non-preemptive, etc...)

We assume stability:

(A1) Fluid limits for work allocations: $\bar{T}_k(t) = \frac{\nu_k}{\mu_k} t$

(A2) $\hat{Q}^n \Rightarrow 0$

(A3) Moment growth rates: $\mathbb{E}[(Q_k(t))^2] = o(t)$ as $t \to \infty$

In multi-class setting: necessary condition: $\sum_{k \in C_i} \frac{\nu_k}{\mu_k} < 1$

Assumption (A1) implies,

$$\lim_{n \to \infty} \bar{D}_{i,j}^n(t) := \bar{D}_{i,j}(t) = \bar{\Phi}_{i,j}(\bar{S}_i(\bar{T}_i(t))) = \nu_{i \to j} t, \text{ u.o.c.}$$
Theorem

$\hat{D}^n$ converges weakly to Brownian Motion with cov matrix,

$$
\Sigma^{(D)} := H \begin{bmatrix}
    \text{diag}(v_k^2) & 0 & \nu_1 \Gamma_1 & \ldots & \nu_K \Gamma_K \\
    0 & \nu_1 \Gamma_1 & \ldots & \nu_K \Gamma_K \\
\end{bmatrix} H'
$$

where $H := \begin{bmatrix}
P_c (I - P')^{-1} & I_{K^2} + P_c (I - P')^{-1} B
\end{bmatrix}$, and

$$
B := 1' \otimes I, \quad P_c := \begin{bmatrix}
P' e_{1,1} \\
P' e_{2,2} \\
\vdots \\
P' e_{K,K}
\end{bmatrix}
$$

Further $\sigma_{i_1 \rightarrow j_1, i_2 \rightarrow j_2} = \Sigma^{(D)}_{(i_1-1)K+j_1, (i_2-1)K+j_2}$
Proof Idea

\[ 0 = \sum_{j=0}^{K} \hat{\Phi}_{i,j}^{n}(\ell), \quad \ell = 1, 2, \ldots \]

\[ \hat{Q}_{k}^{n}(t) = \hat{A}_{k}^{n}(t) + \sum_{j=1}^{K} \hat{D}_{j,k}^{n}(t) - \sum_{j=0}^{K} \hat{D}_{k,j}^{n}(t), \quad t \geq 0 \]
Proof Idea (cont.)

\[ \Phi^*_i(t) := \Phi_i^n(t) \left( \bar{S}^*_i(\bar{T}^*_i(t)) \right), \quad \bar{S}^*_k(t) := \bar{S}^*_k(\bar{T}^*_k(t)) \]

**Lemma:** \( \hat{D}^*_i(t) = \Phi^*_i(t) + p_{i,j} \bar{S}^*_i(t) + p_{i,j} \mu_i \bar{T}^*_i(t) \)

Now manipulate to get:

\[
\hat{D}^*_n(t) = \begin{bmatrix} H & 0_{K \times K} \end{bmatrix} \begin{bmatrix} \hat{A}^*_n(t) \\ \Phi^*_n(t) \\ \bar{S}^*_n(t) \end{bmatrix} - P_c(I - P')^{-1} \hat{Q}^*_n(t).
\]

Now use the FCLT assumptions to get the weak convergence of \( \hat{D}^*_n \)

The (co)variance rate convergence requires a bit more work using the UI and moment assumptions.
Observation: The asymptotic diffusion processes and parameters do not depend on the service sequences

Thought: So how about for a network without delays?

Reason: Queues are stable
The Zero Service Time Model

\[ N_{i,j|k}(\ell) \equiv \text{The number of times that the } \ell\text{'th customer arriving starting at } k \text{ traverses } i \rightarrow j \]

\[ \{ (N_{i,j|k}(\ell), \; i,j \in \{1,\ldots,K\}, \; i \neq j), \; \ell = 1,2,\ldots \} \]

is an i.i.d. sequence (of \( K^2 \) dimensional random vectors) with distribution based on the absorbing Markov chain:

\[ \tilde{P} = \begin{bmatrix} 1 & 0' \\ 1 - P1 & P \end{bmatrix} \]

Now define:

\[ \tilde{D}_{i,j}(t) := \sum_{k=1}^{K} \sum_{\ell=1}^{\ell} N_{i,j|k}(\ell) \]
\[ D_{i,j}(t) \leq \tilde{D}_{i,j}(t), \quad \text{a.s.} \]

Denote now,
\[ \tilde{N}_{i,j}(t) := \tilde{D}_{i,j}(t) - D_{i,j}(t) \]

so,
\[ \tilde{N}_{i,j}(t) = d \sum_{k=1}^{K} \sum_{\ell=1}^{Q_k(t)} N_{i,j|k}(\ell) \]

Hence \( D_{i,j}(t) \) and \( \tilde{D}_{i,j}(t) \) differ by a “stable” quantity

**Proposition:**
\[ \tilde{\sigma}_{i_1 \rightarrow j_1, i_2 \rightarrow j_2} = \sigma_{i_1 \rightarrow j_1, i_2 \rightarrow j_2} \]

**Proposition:**
\[ \tilde{\sigma}_{i_1 \rightarrow j_1, i_2 \rightarrow j_2} = \sum_{k=1}^{K} \alpha_k \text{Cov}(N_{i_1,j_1|k}, N_{i_2,j_2|k}) + \sum_{k=1}^{K} \nu_k^2 \mathbb{E}[N_{i_1,j_1|k} \mathbb{E}[N_{i_2,j_2|k}]] \]
Back to the Example Network

\[ (\alpha_1, v_1) \quad 1 \quad 3 \quad (\alpha_2, v_2) \]

\[ P = \begin{bmatrix}
0 & 0 & 1/2 & 1/2 & 0 & 0 \\
1/2 & 0 & 0 & 1/2 & 0 & 0 \\
1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/2 & 1/2 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix} \]

\[ \mu = \begin{bmatrix}
8.25 \\
8.25 \\
5 \\
8.25 \\
5 \\
5 \\
\end{bmatrix} \]

\[ \alpha = \begin{bmatrix}
1 \\
4 \\
0 \\
0 \\
0 \\
0 \\
\end{bmatrix} \]

\[ V^2 = \begin{bmatrix}
2 \\
2 \\
0 \\
0 \\
0 \\
0 \\
\end{bmatrix} \]
Numerical Example

In this case, $\Sigma^{(D)}$ is $36 \times 36$. The diagonals are:

<table>
<thead>
<tr>
<th>$i \backslash j$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>32/9</td>
<td>20/9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3/2</td>
<td>0</td>
<td>0</td>
<td>3/2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>31/18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>199/18</td>
<td>55/18</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>199/18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$$\rho_{i_1\rightarrow i_2, j_1\rightarrow j_2} := \frac{\sigma_{i_1\rightarrow i_2} \cdot j_1\rightarrow j_2}{\sqrt{\sigma_{i_1\rightarrow i_2}^2 \sigma_{j_1\rightarrow j_2}^2}}.$$ 

E.g. (leaving $v_2$ free):

$$\rho_{1\rightarrow 3,2\rightarrow 4} = \frac{v_2^2 - 4}{\sqrt{(v_2^2 + 4)(v_2^2 + 30)}}.$$
Arrivals to Individual Queues:

$$\Sigma^{(E)} = \begin{bmatrix}
68/9 & 4/3 & 40/9 & 44/9 & 22/9 & 22/9 \\
2 & 2/3 & 10/3 & 5/3 & 5/3 \\
32/9 & 10/9 & 5/9 & 5/9 \\
182/9 & 127/9 & 55/9 \\
199/18 & 55/18 \\
55/18
\end{bmatrix}.$$  

The diagonal of $\Sigma^{(E)}$ may be useful for network decomposition approximations. Normalizing by $\nu$ we get square coefficient of variation:

$$c^2 = \begin{bmatrix}
1.89 & 0.5 & 1.78 & 2.53 & 2.76 & 0.76
\end{bmatrix}'.$$
Single Class (woman brain) vs. Multi-Class (man brain)

Single-class: Each queue has a dedicated (separate) server (all parts of brain can process in parallel)

Multi-class: Queues 1 and 2 are served by the same server under a non-pre-emptive priority policy giving priority to queue 1 (can’t talk while watching football, and priority is obvious). ⇒ Load on server of queues 1 and 2 is $\frac{\nu_1}{\mu_1} + \frac{\nu_2}{\mu_2} \approx 0.97 < 1$
Var(D_{5,4}(t))

1 and 2 on separate servers
1 and 2 on same server
\( \sigma_{5 \to 4}^2 t - \text{Var}(D_{5,4}(t)) \)

1 and 2 on same server

1 and 2 on separate servers
Where does the stochastic brain fit in?

My thoughts are quite random on this, so I don’t know...