QUEUES WITH RESEQUENCING, PART II: HEAVY TRAFFIC LIMITS

by

Subir Varma[1] and Armand M. Makowski[2]

Electrical Engineering Department and Systems Research Center
University of Maryland, College Park, Maryland 20742

ABSTRACT

Queues with resequencing arise as models in various applications, including distributed databases and computer communication networks. Most models are extremely difficult to analyze using traditional techniques. In this paper we investigate the heavy traffic behavior of several resequencing models. It is observed that resequencing delay can be ignored in heavy traffic for a certain class of models, while it blows up in heavy traffic for another class. The heavy traffic limits are combined with light traffic limits in [26] to obtain interpolation approximations.

[1] The work of this author was supported by a fellowship from the Systems Research Center, University of Maryland

[2] The work of this author was supported partially through NSF Grant ECS-83-51836 and partially through NSF Engineering Research Centers Program Grant CDR-88-03012-04

1. Introduction

In this paper our objective is to obtain heavy traffic limits for queues exhibiting the resequencing synchronization constraint. For an introduction to the literature concerning resequencing queues, the reader may consult our companion paper [25] or the survey [3] . This work leads to the following advances in the theory of resequencing systems.

- **(1):** We have obtained good estimates for the queueing delays for several models that were previously intractable analytically;
- **(2):** We have identified a class of models in which resequencing can be ignored in heavy traffic.

We obtain heavy traffic diffusion limits for a variety of resequencing systems possessing the following generic structure: Customers enter a disordering system which they leave (after being served) in an order different from the one in which they entered it. This necessitates resequencing which takes place in a so called resequencing buffer. After leaving the resequencing buffer, the customers enter the buffer of a single server queue from where they leave the system. This generic model is introduced in Section 2, where we also give the recursive equations governing its delays. In Section 3 we obtain the heavy traffic diffusion limit for the generic model from Section 2 from which diffusion limits for specific models can be easily recovered.

In Section 4 we specialize the results of Section 3 for the important special case when the disordering system is an infinite server queue. We show that the queue delay process of this system has the same heavy traffic limit as an ordinary single server queue, i.e., in heavy traffic the resequencing delay has negligible influence on the operation of the system. We also extend this result to the case when there may be more than one disordering and resequencing stages before the single server queue.

In Sections 5 and 6 we obtain the heavy traffic limit for finite server disordering systems. For the case when the disordering system is a $GI/GI/K$ queue, we show that the normalized resequencing delay converges to zero in heavy traffic. For the case when the disordering system is composed of K single server queues operating in parallel, we use an alternate representation for the end–to–end delay of the system than the one given in Section 2, to obtain the heavy traffic diffusion limit. In this case, our results show that the resequencing delay constitutes the major portion of the total delay, in heavy traffic.

2. The model

In this section we introduce a generic resequencing model (Fig 1), from which specific resequencing structures can be recovered as special cases. There is a stream of customers which enter a disordering system, and leave in an order different than the one in which they entered it. After leaving the disordering system, they wait in a resequencing buffer until all customers which entered the disordering system prior to them have left it. After leaving the resequencing box, these customers are served by a single server queue, before finally leaving the system. The model of Baccelli, Gelenbe and Plateau [1] is special case of this model when the disordering system corresponds to an infinite server queue.

We now define some RV's that are useful in discussing the properties of this system. Let the sequences of RV's $\{D_n\}_0^\infty$, $\{v_n\}_0^\infty$ and $\{\tau_n\}_0^\infty$ be defined on some probability space $(\Omega, \mathbb{F}, \mathbb{P})$. Here τ_n represents the time of arrival of the n^{th} customer into the system, D_n represents its disordering delay and v_n represents its service time in the single server queue. In terms of these RV's define the following quantities for all $n = 0, 1 \ldots$,

u_{n+1} : Inter-arrival time between the $(n+1)^{rst}$ and the n^{th} customers $(= \tau_{n+1} - \tau_n)$.

W_n : Delay of the n^{th} customer in the resequencing box and in the buffer of the single server queue.

Y_n : $= D_n + W_n$. This will be referred to as the end-to-end delay in the sequel.

Various kinds of disordering systems can be realized by assuming different statistical structures on the sequence $\{D_n\}_0^\infty$. For example, if the delay sequence $\{D_n\}_0^\infty$ is an iid sequence which is independent of the inter-arrival sequence $\{u_n\}_0^\infty$ then the disordering system corresponds to a $GI/G/\infty$ queue. Similarly we can realize the disordering system as $G/G/K$ queue or a system of K parallel $G/G/1$ queues by imposing a particular structure on $\{D_n\}_0^\infty$.

The analysis of this model is very difficult, one of the reasons for which is that the output stream from the resequencing buffer is a complicated process with batch departures and correlations between the batch sizes and inter–departure times. For the special case when the disordering system has an infinite number of servers, and the the sequences

$\{u_n\}_0^\infty$, $\{D_n\}_0^\infty$ and $\{v_n\}_0^\infty$ are all exponentially distributed, Baccelli, Gelenbe and Plateau [1] were able to derive a complicated expression for the Laplace transform of the end-to-end delay $Y_n = D_n + W_n$. We now derive a recursion first given by Baccelli, Gelenbe and Plateau [1], governing the sequences $\{Y_n\}_0^\infty$ and $\{W_n\}_0^\infty$.

Lemma 2.1 *Consider the resequencing system defined above. If the system is initially empty, then the recursions*

$$Y_0 = D_0$$
$$Y_{n+1} = \max\{D_{n+1}, Y_n + v_n - u_{n+1}\}, \qquad n = 0, 1 \ldots (2.1)$$

and

$$W_0 = 0$$
$$W_{n+1} = \max\{0, W_n + D_n - D_{n+1} + v_n - u_{n+1}\}, \qquad n = 0, 1 \ldots (2.2)$$

hold.

For a proof, the reader may consult [1] or [25].

We shall assume that

(Ia): The sequences $\{u_n\}_0^\infty$ and $\{v_n\}_0^\infty$ are iid with finite second moments and mutually independent.

For all $n = 0, 1 \ldots$, we set

$$u = \mathbb{E}(u_{n+1}) < \infty, \quad \sigma_U^2 = Var(u_{n+1}) < \infty$$

and

$$v = \mathbb{E}(v_n) < \infty, \quad \sigma_V^2 = Var(v_n) < \infty.$$

3. The heavy traffic limit for general resequencing systems

In this section we obtain heavy traffic diffusion limits for resequencing systems possessing a disordering system which can have an arbitrary structure. Disordering systems possessing specific structures are discussed in Sections 4–6.

We now consider a sequence of resequencing systems indexed by $r = 1, 2\ldots$, each of which satisfies assumptions **(Ia)**. Moreover assume that

(Ib): As $r \uparrow \infty$,

$$\sigma_U(r) \to \sigma_U,$$

$$\sigma_V(r) \to \sigma_V,$$

$$[u(r) - v(r)]\sqrt{r} \to c.$$

(Ic): For some $\epsilon > 0$,

$$\sup_r \{\mathbb{E}\{\mid u_1(r) \mid^{2+\epsilon}\}, \mathbb{E}\{v_1(r) \mid^{2+\epsilon}\} < \infty.$$

For $r = 1, 2\ldots$, define the following partial sums

$$V_0(r) = 0,$$

$$V_n(r) = v_0(r) + \ldots + v_{n-1}(r), \qquad n = 1, 2\ldots \tag{3.1}$$

and

$$U_0(r) = 0,$$

$$U_n(r) = u_0(r) + \ldots + u_{n-1}(r). \qquad n = 1, 2\ldots \tag{3.2}$$

For $r = 1, 2\ldots$ define the stochastic processes $\xi^j \equiv \{\xi_t^j(r), t \geq 0\}, j = 0, 1$, with sample paths in $D[0, \infty)$ by

$$\xi_t^0(r) = \frac{U_{[rt]}(r) - u(r)[rt]}{\sqrt{r}}, \quad t \geq 0 \tag{3.3}$$

and

$$\xi_t^1(r) = \frac{V_{[rt]}(r) - v(r)[rt]}{\sqrt{r}}, \quad t \geq 0. \tag{3.4}$$

Let $\xi^j \equiv \{\xi_t^j, t \geq 0\}, j = 0, 1$, be two independent Wiener processes. Lemma 3.1 shows that the stochastic processes defined in (3.1)-(3.2) converge weakly to these Wiener processes.

Lemma 3.1. *As* $r \uparrow \infty$,

$$(\xi^0(r), \xi^1(r)) \Rightarrow (\sigma_U \xi^0, \sigma_V \xi^1) \tag{3.5}$$

in $D[0,\infty)^2$.

Proof. Equation (3.5) follows directly by Prohorov's functional central limit theorem for triangular arrays [22] under assumptions **(Ia)-(Ic)**. ∎

For $r = 1, 2 \ldots$, we set

$$S_0(r) = 0$$
$$S_n(r) = V_n(r) - U_n(r), \qquad n = 1, 2 \ldots (3.6)$$

and define the stochastic processes $\zeta \equiv \{\zeta_t(r), t \geq 0\}$, with sample paths in $D[0,\infty)$, by

$$\zeta_t(r) = \frac{S_{[rt]}(r)}{\sqrt{r}}, \quad t \geq 0. \tag{3.7}$$

We also define the stochastic process $\zeta \equiv \{\zeta_t, t \geq 0\}$, by

$$\zeta_t = \sigma_V \xi_t^1 - \sigma_U \xi_t^0 - ct, \quad t \geq 0. \tag{3.8}$$

Lemma 3.2 shows that the stochastic processes generated by the random walk process (3.7) converge to ζ in the limit.

Lemma 3.2. *As* $r \uparrow \infty$,

$$\zeta(r) \Rightarrow \zeta \tag{3.9}$$

in $D[0,\infty)$.

Proof. Fix $r \geq 1$ and $t \geq 0$. We see from (3.6) that

$$\begin{aligned}
\zeta_t(r) &= \frac{V_{[rt]}(r) - U_{[rt]}}{\sqrt{r}} \\
&= \frac{V_{[rt]}(r) - v(r)[rt]}{\sqrt{r}} - \frac{U_{[rt]}(r) - u(r)[rt]}{\sqrt{r}} - \frac{[rt][u(r) - v(r)]}{\sqrt{r}} \\
&= \xi_t^1(r) - \xi_t^0(r) - \frac{[rt]}{r}[u(r) - v(r)]\sqrt{r}
\end{aligned}$$

From assumption **(IIb)** it is clear that as $r \uparrow \infty$,

$$\frac{[rt]}{r}[u(r) - v(r)]\sqrt{r} \to ct,$$

and we conclude to (3.9) by invoking Lemma 2.2.1 and the continuous mapping theorem [4, Theorem 5.1].

∎

For $r = 1, 2 \ldots$, we define the stochastic process $\mu \equiv \{\mu_t(r), t \geq 0\}$ and $\delta \equiv \{\delta_t(r), t \geq 0\}$ with sample paths in $D[0,\infty)$ by

$$\mu_t(r) = \frac{W_{[rt]}(r)}{\sqrt{r}}, \quad t \geq 0 \tag{3.10}$$

and

$$\delta_t(r) = \frac{D_{[rt]}(r)}{\sqrt{r}}, \quad t \geq 0. \tag{3.11}$$

Theorem 3.1.

(a): *Assume that as $r \uparrow \infty$,*

$$D_0(r) \xrightarrow{\mathcal{D}} D_0 \tag{3.12a}$$

and

$$(\delta(r), \zeta(r)) \Rightarrow (\delta, \zeta) \tag{3.12b}$$

in $D[0,\infty)^2$. Further assume that $\mid c \mid < \infty$, then

$$\mu(r) \Rightarrow g(\zeta - \delta) \tag{3.13}$$

in $D[0,\infty)$ as $r \uparrow \infty$.

(b): *Assume that as $r \uparrow \infty$,*

$$(\delta(r), \xi^0(r), \xi^1(r)) \Rightarrow (\delta, \xi^0, \xi^1) \tag{3.14a}$$

in $D[0,\infty)^3$ and

$$u(r) \to u \text{ and } v(r) \to v \text{ with } u(r) > v(r) \text{ and } u > v, \tag{3.14b}$$

then

$$\mu(r) \Rightarrow 0 \tag{3.15}$$

in $D[0,\infty)$ *as* $r \uparrow \infty$.

Proof. We first prove Part (a). Fix $r = 1, 2\ldots.$ We can write the recursion (2.2) for the waiting time sequence as

$$W_0(r) = 0,$$
$$W_{n+1}(r) = \max\{0, W_n(r) + X_{n+1}(r)\}, \qquad n = 0, 1, \ldots \quad (3.16)$$

where

$$X_{n+1}(r) = D_n(r) - D_{n+1}(r) + v_n(r) - u_{n+1}(r). \qquad n = 0, 1 \ldots \quad (3.17)$$

By successive substitutions, we obtain

$$W_n(r) = \max\{0, X_n(r), X_n(r) + X_{n-1}(r), \ldots, X_n(r) + \ldots + X_1(r)\}. \quad n = 0, 1 \ldots \quad (3.18)$$

Let

$$Z_0(r) = 0,$$
$$Z_n(r) = \sum_{i=1}^{n} X_i(r). \qquad n = 1, 2 \ldots \quad (3.19)$$

It follows that

$$W_n(r) = Z_n(r) - \min_{0 \le k \le n} Z_k(r). \qquad n = 0, 1 \ldots \quad (3.20)$$

Note that

$$Z_n(r) = D_0(r) - D_n(r) + S_n(r). \qquad n = 0, 1 \ldots \quad (3.21)$$

For $r = 1, 2 \ldots$ we introduce the stochastic process $\rho(r) \equiv \{\rho_t(r), t \ge 0\}$ with sample paths in $D[0,\infty)$ by

$$\rho_t(r) = \frac{Z_{[rt]}(r)}{\sqrt{r}}, \quad t \ge 0. \quad (3.22)$$

From (3.20) and (3.22) it follows that

$$\mu_t(r) = g(\rho(r))_t, \quad t \ge 0. \quad (3.23)$$

Hence by the continuous mapping theorem, in order to prove (3.13), it is sufficient to show that as $r \uparrow \infty$,

$$\rho(r) \Rightarrow \zeta - \delta \tag{3.24}$$

in $D[0, \infty)$. From (3.21), we see that

$$\rho(r) = \frac{D_0(r)}{\sqrt{r}} + \zeta(r) - \delta(r). \tag{3.25}$$

As a consequence of (3.12a), it follows that

$$\frac{D_0(r)}{\sqrt{r}} \Rightarrow 0 \text{ as } r \uparrow \infty$$

so that (3.24) follows from (3.12b), (3.25) and the converging together theorem.

We now provide a proof for Part (b), the main idea of which is borrowed from Iglehart and Whitt [13]. For $r = 1, 2 \ldots$ introduce the stochastic processes $\rho'(r) \equiv \{\rho'_t(r), t \geq 0\}$, with sample paths in $D[0, \infty)$, by

$$\rho'_t(r) = \frac{Z_{[rt]}(r) - [v(r) - u(r)][rt]}{\sqrt{r}}, \quad t \geq 0. \tag{3.26}$$

Then proceeding as in Part (a), it can be easily shown with the help of (3.14a) that

$$\rho'(r) \Rightarrow \sigma_V \xi^1 - \sigma_U \xi^0 - \delta \tag{3.27}$$

in $D[0, \infty)$ as $r \uparrow \infty$.

In order to prove that $\mu(r) \Rightarrow 0$, it is sufficient to show for each $T > 0$, that

$$\sup_{0 \leq t \leq T} | \mu_t(r) | \xrightarrow{\mathbb{P}} 0 \tag{3.28}$$

as $r \uparrow \infty$. It is intuitive to expect that (3.28) would be true, since

$$\mu_t(r) = \rho_t(r) - \inf_{0 \leq s \leq t} \rho_s(r), \quad t \geq 0$$

and as a consequence of (3.14b)

$$\lim_{r\uparrow\infty}[u(r)-v(r)]\sqrt{r}=\infty$$

so that $\rho_t(r)\downarrow -\infty$ as $r\uparrow\infty$.

Fix $T>0$, a value d in $[0,T]$ and $0<\epsilon<1$. We first show that as $r\uparrow\infty$, with probability greater that $1-\epsilon$, we have

$$\inf_{0\le s\le t-d}\frac{S_{[rs]}(r)-D_{[rs]}(r)}{\sqrt{r}}\ge\frac{S_{[rt]}(r)-D_{[rt]}(r)}{\sqrt{r}},\quad 0\le t\le T \tag{3.29}$$

which is a justification for the intuitive fact that $\inf_{0\le s\le t}\rho_s(r)\approx\rho_t(r)$ for a sufficiently large value of r. For $d\le t\le T$, we note that

$$\inf_{0\le s\le t-d}\frac{S_{[rs]}(r)-D_{[rs]}(r)}{\sqrt{r}}$$

$$\begin{aligned}
&=\inf_{0\le s\le t-d}\left(\frac{S_{[rs]}(r)-D_{[rs]}(r)}{\sqrt{r}}-\frac{[v(r)-u(r)][rs]}{\sqrt{r}}+\frac{[v(r)-u(r)][rs]}{\sqrt{r}}\right)\\
&\ge\inf_{0\le s\le t-d}\left(\frac{S_{[rs]}(r)-D_{[rs]}(r)}{\sqrt{r}}-\frac{[v(r)-u(r)][rs]}{\sqrt{r}}\right)\\
&\quad+\frac{[v(r)-u(r)][r(t-d)]}{\sqrt{r}}\\
&=\inf_{0\le s\le t-d}\left(\frac{S_{[rs]}(r)-D_{[rs]}(r)}{\sqrt{r}}-\frac{[v(r)-u(r)][rs]}{\sqrt{r}}\right)\\
&\quad-\left(\frac{S_{[rt]}(r)-D_{[rt]}(r)}{\sqrt{r}}-\frac{[v(r)-u(r)][rt]}{\sqrt{r}}\right)\\
&\quad+\frac{S_{[rt]}(r)-D_{[rt]}(r)}{\sqrt{r}}+\frac{[u(r)-v(r)][rd]}{\sqrt{r}}\\
&\ge\frac{S_{[rt]}(r)-D_{[rt]}(r)}{\sqrt{r}}
\end{aligned}$$

with probability greater that $1-\epsilon$ for sufficiently large $r\ge r_0$. The second inequality follows from assumption (3.14b), while the last inequality follows from the fact that the terms in the first two brackets have weak limits while the last term blows to infinity as r increases.

As a result of (3.29), it follows that for $r \geq r_0$, with probability greater than $1 - \epsilon$, we have

$$\inf_{0 \leq s \leq t} \frac{S_{[rs]}(r) - D_{[rs]}(r)}{\sqrt{r}} = \inf_{t-d \leq s \leq t} \frac{S_{[rs]}(r) - D_{[rs]}(r)}{\sqrt{r}}$$

for a fixed value of d. Hence, for $r \geq r_0$, with probability greater that $1 - \epsilon$, we see that

$$\begin{aligned} \sup_{0 \leq t \leq T} \mid \mu_t(r) \mid & \\ &= \sup_{0 \leq t \leq T} \mid \frac{S_{[rt]}(r) - D_{[rt]}(r)}{\sqrt{r}} - \inf_{t-d \leq s \leq t} \frac{S_{[rs]}(r) - D_{[rs]}(r)}{\sqrt{r}} \mid \\ &\leq \sup_{0 \leq s,t \leq T} \sup_{|s-t|<d} \mid \frac{Z_{[rt]}(r) - Z_{[rs]}(r)}{\sqrt{r}} - \frac{[v(r) - u(r)][r(t-s)]}{\sqrt{r}} \mid \\ &= \omega_{\rho'(r)}(d) \end{aligned} \tag{3.30}$$

where the modulus of continuity $\omega_{\rho'(r)}$ for the process $\rho'(r)$ is given by

$$\omega_{\rho'(r)}(d) = \sup_{0 \leq s,t \leq T} \sup_{|s-t| \leq d} \mid \rho'_t(r) - \rho'_s(r) \mid, \quad d > 0.$$

Also as a result of (3.27) and of the continous mapping theorem, we get

$$\omega_{\rho'(r)}(d) \Rightarrow \omega_{\rho'}(d) \tag{3.31}$$

as $r \uparrow \infty$. Since we can make the value of d as small as we please, and since

$$\omega_{\rho'}(d) \xrightarrow{\mathbb{P}} 0 \tag{3.32}$$

as $d \downarrow 0$, it follows from (3.30) that

$$\mu(r) \Rightarrow 0 \tag{3.33}$$

in $D[0, T]$ as $r \uparrow \infty$, and this proves the theorem. ∎

Part (b) of Theorem 3.1 implies the surprising fact that the normalized resequencing delay of the customers will always be zero if the single server queue is operating in its stable regime, irrespective of whether the disordering system is in heavy traffic or not. However this result hinges upon the crucial condition (3.14), that $\delta_r \Rightarrow \delta$ in $D[0,\infty)$ as $r \uparrow \infty$. This condition is satisfied for disordering systems of the $GI/GI/K$ type as well as for infinite server disordering systems. Unfortunately it is not satisfied for disordering systems within which there is probabilistic routing of customers. For example, as elaborated in Section 6, this condition is not satisfied for disordering systems which are made up of parallel queues with Bernoulli switching of arriving customers, or disordering systems which involve probabilistic feedback from the output to the input. For such disordering systems, the conclusion of Part (b) of the Theorem does not hold. In fact we show in Section 6, that rather then going to zero, the resequencing delay constitutes the major portion of the total delay of such systems, in heavy traffic.

4. The heavy traffic limit for infinite server systems

In this section we specialize the results of Theorem 3.1 to the case when the disordering system is an infinite server queue. We shall assume that

(Id): The sequences $\{u_n\}_0^\infty$, $\{D_n\}_0^\infty$ and $\{v_n\}_0^\infty$ are iid with finite second moments and independent.

For all $n = 0, 1 \ldots$, we set

$$u = \mathbb{E}(u_n) < \infty, \quad \sigma_U^2 = Var(u_n) < \infty,$$

$$v = \mathbb{E}(v_n) < \infty, \quad \sigma_V^2 = Var(v_n) < \infty$$

and

$$d = \mathbb{E}(D_n) < \infty, \quad \sigma_D^2 = Var(D_n) < \infty.$$

We now consider a sequence of resequencing systems indexed by $r = 1, 2 \ldots$, each of which satisfies assumption **(Id)**. Moreover assume that

(Ie): As $r \uparrow \infty$,

$$\sigma_U(r) \to \sigma_U,$$

$$\sigma_V(r) \to \sigma_V,$$

$$\sigma_D(r) \to \sigma_D,$$

$$[u(r) - v(r)]\sqrt{r} \to c.$$

(If): For some $\epsilon > 0$,

$$\sup_r \{\mathbb{E}\{| u_1(r) |^{2+\epsilon}\}, \mathbb{E}\{v_1(r) |^{2+\epsilon} \mathbb{E}\{| D_1(r) |^{2+\epsilon}\}\} < \infty$$

Under these assumptions, Theorem 3.1 immediately yields the following corollary.

Corollary 4.1. *As* $r \uparrow \infty$

$$\mu(r) \Rightarrow g(\zeta) \tag{4.1}$$

in $D[0, \infty)$.

Proof. This follows from Part (a) of Theorem 3.1 owing to the fact that

$$\delta_t(r) = \frac{D_{[rt]}(r)}{\sqrt{r}} \Rightarrow 0$$

in $D[0, \infty)$ as $r \uparrow \infty$. ∎

For infinite server disordering systems the sequence $\{W_n\}_0^\infty$ has the same traffic limit as the sequence of waiting times in an ordinarysingle server queue. This means that asymptotically resequencing as a negligible effect on the operation of the single server queue in heavy traffic. This result is surprising if seen from the following viewpoint: Kingman [15] has shown that the diffusion limit for a single server queue depends on the particular discipline chosen to serve the customers, for example it is different if the customers are served in LCFS order rather than in FCFS order. Resequencing may be viewed as a special type of service discipline (if the resequencing buffer and the single server queue buffer are regarded as a single buffer of an equivalent single server), because customers are served in the order in which they entered the infinite server queue, rather than the order in which they enter the equivalent buffer. Also note that this effect remains unchanged in heavy traffic. Hence in this case even though we change the service discipline of the single server queue, we nevertheless obtain the same diffusion limit.

4.1 Generalization to a tandem system

We now proceed to extend the result of the previous sub–section to the case when there are an arbitrary number of disordering and resequencing systems preceding the single server queue. The system under consideration operates as follows: Each customer is disordered by an infinite server queue and resequenced K successive times before it enters the buffer of a single server queue. After getting served there, it leaves the system (Fig 2). Let the sequences $\{u_n\}_0^\infty$ and $\{v_n\}_0^\infty$ be defined as before, and for each $1 \leq k \leq K$ and $n = 0, 1 \ldots$, define the following,

D_n^k : Delay of the n^{th} customer at the k^{th} disordering system;

W_n^k : For $1 \leq k \leq K-1$, this RV represents the delay of the n^{th} customer in the k^{th} resequencing box;

W_n : Delay of the n^{th} customer in the K^{th} resequencing box plus the delay in the buffer of the single server queue; and

D_n : $= D_n^1 + W_n^1 + \ldots + D_n^{K-1} + W_n^{K-1} + D_n^K$, i.e., the total disordering delay of the n^{th} customer, before it is resequenced and sent to the buffer of the single server queue.We shall assume that

(Ig): The sequences $\{v_n\}_0^\infty$, $\{u_n\}_0^\infty$ and $\{D_n^k\}_0^\infty, 1 \leq k \leq K$, are iid with finite second moments and mutually independent.

For all $n = 0, 1 \ldots$, we set

$$u = \mathbb{E}(u_n) < \infty, \quad \sigma_U^2 = Var(u_n) < \infty$$

$$v = \mathbb{E}(v_n) < \infty, \quad \sigma_V^2 = Var(v_n) < \infty$$

and

$$d_k = \mathbb{E}(D_n^k) < \infty, \quad \sigma_k^2 = Var(D_n^k) < \infty, \quad 1 \leq k \leq K.$$

Now consider a sequence of resequencing systems indexed by $r = 1, 2 \ldots$ each of which satisfies assumptions **(Ig)**. Moreover assume that

(Ih): As $r \uparrow \infty$,

$$\sigma_U(r) \to \sigma_U,$$

$$\sigma_V(r) \to \sigma_V,$$

$$\sigma_k(r) \to \sigma_k, \quad 1 \leq k \leq K$$

$$[u(r) - v(r)]\sqrt{r} \to c.$$

(Ii): For some $\epsilon > 0$,

$$\sup_{r,k} \{ I\!E\{ | u_1(r) |^{2+\epsilon}\}, I\!E\{v_1(r) |^{2+\epsilon}\} I\!E\{| D_1^k(r) |^{2+\epsilon}\}\} < \infty$$

Define the partial sums $\{V_n\}_0^\infty$, $\{U_n\}_0^\infty$, $\{X_n\}_0^\infty$ and $\{Z_n\}_0^\infty$ as in (3.1), (3.2), (3.17) and (3.19) respectively. Also define the stochastic processes $\xi^0(r), \xi^1(r)$, $\zeta(r)$, $\mu(r)$, $\delta(r)$ and $\rho(r)$ as in (3.1), (3.4), (3.7), (3.10), (3.11) and (3.22). It is easy to see that Lemma 7.3.1 and Lemma 7.3.2 continue to hold for this model.

For $r = 1, 2\ldots$, define the stochastic processes $\mu^k \equiv \{\mu_t^k, t \geq 0\}, 1 \leq k \leq K-1$ with sample paths in $D[0, \infty)$ by

$$\mu_t^k(r) = \frac{W_{[rt]}^k(r)}{\sqrt{r}}, \quad 1 \leq k \leq K-1. \tag{4.2}$$

Theorem 4.1. *As* $r \uparrow \infty$

$$\delta(r) \Rightarrow 0 \tag{4.3}$$

in $D[0, \infty)$.

Proof. Fix $r = 1, 2 \ldots$. Note that

$$\delta_t(r) = \frac{D_{[rt]}^1}{\sqrt{r}} + \mu_t^1(r) + \ldots + \frac{D_{[rt]}^{K-1}}{\sqrt{r}} + \mu_t^{K-1}(r) + \frac{D_{[rt]}^K}{\sqrt{r}}, \quad t \geq 0. \tag{4.4}$$

Hence in order to prove (4.3), it is sufficient to show that as $r \uparrow \infty$

$$(\mu^1(r), \ldots, \mu^{K-1}(r)) \Rightarrow (0, \ldots, 0) \tag{4.5}$$

in $D[0, \infty)^{K-1}$.

We propose to prove (4.5) by induction on the number of levels in the system. We first show that as $r \uparrow \infty$

$$\mu^1(r) \Rightarrow 0 \tag{4.6}$$

in $D[0, \infty)$.

Note that

$$W^1_{n+1}(r) = \max\{0, W^1_n(r) + D^1_n(r) - D^1_{n+1}(r) - -u_{n+1}(r)\}$$
$$= Z^1_{n+1}(r) - \min_{0\le i\le n+1} Z^1_i(r), \qquad n = 0, 1 \ldots (4.7)$$

where

$$Z^1_n(r) = D^1_0(r) - D^1_n(r) - u_1(r) - \ldots - u_n(r). \qquad n = 0, 1 \ldots (4.8)$$

From (4.7)–(4.8) it follows that

$$\mu^1_t(r) = \frac{D^1_0(r)}{\sqrt{r}} - \frac{D^1_{[rt]}(r)}{\sqrt{r}} - \frac{U_{[rt]}(r)}{\sqrt{r}} - \inf_{0\le s\le t}\{\frac{D^1_0(r)}{\sqrt{r}} - \frac{D^1_{[rs]}(r)}{\sqrt{r}} - \frac{U_{[rs]}(r)}{\sqrt{r}}\}, \quad t \ge 0 \quad (4.9)$$

For $r = 1, 2 \ldots$, define the stochastic processes $\hat{\mu}^1(r) \equiv \{\hat{\mu}^1_t(r), t \ge 0\}$ in $D[0, \infty)$ by

$$\hat{\mu}^1_t(r) = -\frac{U_{[rt]}(r)}{\sqrt{r}} - \inf_{0\le s\le t}\{-\frac{U_{[rs]}(r)}{\sqrt{r}}\}, \quad t \ge 0 \quad (4.10)$$

and note that $\mu^1(r)$ and $\hat{\mu}^1(r)$ both have the same limit due to the converging together theorem. Note that in this case we do not require an additional condition such as (3.12a) to conclude that $\frac{D^1_0(r)}{\sqrt{r}}$ or $\frac{D^1_{[rt]}(r)}{\sqrt{r}}$ converges to zero, since by assumption they form sequences ofiid RVs.

Hence in order to prove (4.6), it is sufficient to show that $\hat{\mu}^1(r) \Rightarrow 0$ as $r \uparrow \infty$. In order to prove this, it is sufficient to show that

$$\sup_{0\le t\le 1} \mid \hat{\mu}^1_t(r) \mid \xrightarrow{\mathbb{P}} 0 \quad (4.11)$$

as $r \uparrow \infty$. The proof for (4.11) is similar to the proof given for Part (b) of Theorem 7.3.1, and is therefore omitted.

As the induction step, assume that as $r \uparrow \infty$

$$(\mu^1(r), \ldots, \mu^k(r)) \Rightarrow (0, \ldots, 0) \quad (4.12)$$

in $D[0, \infty)^k$, for some $2 \le k \le K - 2$. We shall show that as $r \uparrow \infty$

$$(\mu^1(r), \ldots, \mu^{k+1}(r)) \Rightarrow (0, \ldots, 0) \quad (4.13)$$

in $D[0,\infty)^{k+1}$.

For $1 \le k \le K-1$ and $r = 1, 2 \ldots$, define the RVs $T_n^k(r)$ by

$$T_n^k(r) = D_n^1(r) + W_n^1(r) + \ldots + D_n^k(r) + W_n^k(r)$$

Note that as a consequence of (4.12), it follows that as $r \uparrow \infty$,

$$\frac{T_{[r.]}^k(r)}{\sqrt{r}} \Rightarrow 0 \tag{4.14}$$

in $D[0,\infty)$.

Note that

$$\begin{aligned} W_{n+1}^{k+1}(r) &= \max\{0, W_n^{k+1}(r) + T_n^k(r) + D_n^{k+1}(r) - T_{n+1}^k(r) - D_{n+1}^{k+1}(r) - u_{n+1}(r)\} \\ &= Z_{n+1}^{k+1}(r) - \min_{0 \le i \le n+1} Z_i^{k+1}(r), \qquad n = 0, 1 \ldots \end{aligned} \tag{4.15}$$

where

$$Z_n^{k+1}(r) = T_0^k(r) + D_0^{k+1}(r) - T_n^k(r) - D_n^{k+1}(r) - u_1(r) - \ldots - u_n(r) \quad n = 0, 1 \ldots \tag{4.16}$$

From (4.15)–(4.16) it follows that

$$\begin{aligned} \mu_t^{k+1}(r) =& \frac{T_0^1(r)}{\sqrt{r}} + \frac{D_0^{k+1}(r)}{\sqrt{r}} - \frac{T_{[rt]}^k(r)}{\sqrt{r}} - \frac{D_{[rt]}^{k+1}(r)}{\sqrt{r}} - \frac{U_{[rt]}(r)}{\sqrt{r}} \\ & - \inf_{0 \le s \le t}\{\frac{T_0^1(r)}{\sqrt{r}} + \frac{D_0^{k+1}(r)}{\sqrt{r}} - \frac{T_{[rs]}^k(r)}{\sqrt{r}} - \frac{D_{[rs]}^{k+1}(r)}{\sqrt{r}} - \frac{U_{[rs]}(r)}{\sqrt{r}}\} \end{aligned} \tag{4.17}$$

From equation (4.14) (which is a consequence of the induction hypothesis (4.12)) and (4.17), and the fact that $T_n^k(r)$ is independent of $D_n^{k+1}(r)$, it is clear that that (4.13) holds, and this completes the proof.

∎

Equation (4.3) in combination with Part (a) of Theorem 3.1, implies that

Theorem 4.2. *As* $r \uparrow \infty$

$$\mu(r) \Rightarrow g(\zeta) \tag{4.18}$$

in $D[0,\infty)$.

5. The heavy traffic limit for finite server resequencing systems: Multiserver disordering systems

The resequencing systems to be analyzed in the next two sections deviate slightly from the general model introduced in Section 2 due to the fact that the single server queue after the resequencing buffer is omitted from the system. In this section we shall consider the case when the disordering system is a $GI/GI/K$ queue, while in Section 6 we shall consider the case when the disordering system consists of K single server queues operating in parallel.

The resequencing system under consideration operates as follows: Customers enter a $GI/GI/K$ queue, after obtaining service from which they are resequenced in a resequencing buffer and leave the system. Our basic heavy traffic result about this system is stated next.

Theorem 5.1 *The end–to–end delay in the $GI/GI/K$ resequencing system has the same heavy traffic limit as the response time of a $GI/GI/K$ queue.*

Proof. This system satisfies the conditions in Part (b) of Theorem 3.1, so that the conclusion is a direct consequence of (3.15). ∎

Let the average end–to–end delay for the system be denoted by $\overline{T}_K(\lambda)$. Then Theorem 5.1 and results regarding heavy traffic limits for $GI/G/K$ queues in Kollerstrom [16] imply that

$$\lim_{\lambda\uparrow K\mu}(K\mu-\lambda)\overline{T}_K(\lambda)=[\sigma_U^2+\frac{\sigma_V^2}{K^2}]\frac{K^2\mu^2}{2} \qquad K=2,3\ldots \quad (5.1)$$

where λ,σ_U and μ,σ_V are the rates and variances of the arrival and service processes respectively.

6. The heavy traffic limit for finite server resequencing systems: Disordering due to parallel queues

The system to be analyzed has a disordering system composed of K parallel single server queues (Fig 3). We assume that customers are switched to the different queues by a Bernoulli switch with switching probability $p_k, 1 \leq k \leq K$. After the customers receive service, they are resequenced in a resequencing buffer after which they leave the system.

This system was analyzed by Gün and Jean-Marie [16] for the special case when the arrival process into the system is Poissonian. They gave a complicated expression for the average end–to–end delay involving the virtual waiting time in the system. However, since for most systems it is difficult to obtain a formula for the virtual waiting time, we expect that the limit theorem approximations that we obtain to be of practical computational value.

The following RV's are defined on a common probability space $(\Omega, \mathbb{F}, \mathbb{P})$. For $n = 0, 1 \ldots$ and $1 \leq k \leq K$,

u_{n+1} : Inter–arrival time between the $(n+1)^{rst}$ and n^{th} customers.

v_n^k : Service time of the n^{th} customer to enter the system, if it were to join the k^{th} queue.

a_n^k : This is a $\{0, 1\}$–valued RV, such that $a_n^k = 1$ implies that the n^{th} customer joins the k^{th} queue.

W_n^k : Waiting time of the n^{th} customer to enter the system, if it were to join the k^{th} queue.

T_k : End–to–end delay of the n^{th} customer to enter the system (including the resequencing delay).

We shall assume that

(Ij): The sequences $\{u_{n+1}\}_0^\infty$, $\{a_n^k\}_0^\infty$ and $\{v_n^k\}_0^\infty, 1 \leq k \leq K$, are iid with finite second moments, and mutually independent.

For $n = 0, 1 \ldots$, we set

$$\mathbb{P}(a_n^k) = p_k,$$

$$u = \mathbb{E}(u_{n+1}) < \infty, \quad \sigma_0^2 = Var(u_{n+1}) < \infty$$

and

$$v_k = \mathbb{E}(v_n^k) < \infty, \quad \sigma_k^2 = Var(v_n^k) < \infty, \quad 1 \leq k \leq K$$

6.1 Recursive representation for the delays

The delays in the system obey the recursions given in Lemma 2.1 with D_n replaced by the response time of the n^{th} customer in the system of parallel queues. However we give another set of recursions for the system which have the advantage of facilitating the proof of the heavy traffic limit theorems. Assuming that the initial batch arrives into an empty system at time $t = 0$, it is easy to see that for each $1 \leq k \leq K$,

$$W_0^k = 0$$
$$W_{n+1}^k = \max\{0, W_n^k + a_n^k v_n^k - u_{n+1}\}. \qquad n = 0, 1 \ldots (6.1)$$

The end–to–end delay $T_n, 1 \leq k \leq K$, is given by

$$T_n = \max_{1 \leq k \leq K} \{W_n^k + a_n^k \sigma_n^k\}. \qquad n = 0, 1 \ldots (6.2)$$

It is well known [1] that the stability condition of a system with resequencing is the same as the system without resequencing. Therefore the system is stable iff each queue is stable, i.e.,

$$p_k v_k < u, \quad 1 \leq k \leq K. \tag{6.3}$$

6.2 The diffusion limit

We now proceed with the task of obtaining heavy traffic diffusion limits for the delay processes in the resequencing system. We consider a sequence of resequencing systems indexed by $r = 1, 2 \ldots$, each of which satisfies assumption **(Ij)**. We make the following additional assumptions **(Ik)**–**(Il)**, where

(Ik): As $r \uparrow \infty$,

$$\sigma_k(r) \to \sigma_k, \quad 0 \leq k \leq K,$$
$$p_k(r) \to p_k, \quad 1 \leq k \leq K,$$
$$v_k(r) \to v_k, \quad 1 \leq k \leq K,$$
$$[u(r) - p_k(r) v_k(r)]\sqrt{r} \to c_k, \quad 1 \leq k \leq K.$$

(Il): For some $\epsilon > 0$,

$$\sup_{r,k} \{\mathbb{E}\{| u_1(r) |^{2+\epsilon}\}, \mathbb{E}\{| v_1^k(r) |^{2+\epsilon}\}\} < \infty.$$

For $r = 1, 2 \ldots$, define the following partial sums

$$V_0^k(r) = 0,$$
$$V_n^k(r) = a_0^k v_0^k(r) + \ldots + a_{n-1}^k v_{n-1}^k(r), \quad 1 \le k \le K, \quad n = 1, 2 \ldots (6.4)$$
$$U_0(r) = 0,$$
$$U_n(r) = u_1(r) + \ldots + u_n(r). \qquad n = 1, 2 \ldots (6.5)$$

and

$$S_0^k(r) = 0$$
$$S_n^k(r) = V_n^k(r) - U_n(r), \ 1 \le k \le K \qquad n = 1, 2 \ldots (6.6)$$

and define the stochastic processes $\zeta^k(r) \equiv \{\zeta_t^k(r), t \ge 0\}, 1 \le k \le K$, with sample paths in $D[0, \infty)$ by

$$\zeta_t^k(r) = \frac{S_{[rt]}^k(r)}{\sqrt{r}}, \quad 1 \le k \le K, \quad t \ge 0. \tag{6.7}$$

Let $\xi \equiv \{\xi_t^k, t \ge 0\}, 1 \le k \le K$, be K independent Wiener processes and define the stochastic processes $\zeta^k \equiv \{\zeta_t^k, t \ge 0\}, 1 \le k \le K$, by

$$\zeta_t^k = \sum_{j=1}^{K} Q_{kj} \xi_t^j - c_k t, \quad 1 \le k \le K, \quad t \ge 0 \tag{6.8}$$

where the matrix $Q \equiv \{Q_{ij}\}_{i,j=1}^K$ is such that the covariance matrix R for the diffusion is given by (with $p_k + \overline{p}_k = 1, 1 \le k \le K$),

$$R = QQ^T$$

$$= \begin{pmatrix} \sigma_0^2 + p_1\sigma_1^2 + p_1\overline{p}_1 v_1^2 & \sigma_0^2 - p_1 p_2 v_1 v_2 & \cdots & \sigma_0^2 - p_1 p_K v_1 v_K \\ \sigma_0^2 - p_2 p_1 v_2 v_1 & \sigma_0^2 + p_2\sigma_2^2 + p_2\overline{p}_2 v_2^2 & \cdots & \sigma_0^2 - p_2 p_K v_2 v_K \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_0^2 - p_K p_1 v_K v_1 & \sigma_0^2 - p_K p_2 v_K v_2 & \cdots & \sigma_0^2 + p_K\sigma_K^2 + p_K\overline{p}_K v_K^2 \end{pmatrix}. \tag{6.9}$$

The process $(\zeta^1, \ldots, \zeta^K)$ is thus a K–dimensional diffusion process with drift vector $c = (-c_1, \ldots, -c_K)$ and covariance matrix R.

Theorem 6.2 shows that the stochastic processes (6.7) generated by the random walk (6.6) converge to $(\zeta^1, \ldots, \zeta^K)$ in the limit.

Theorem 6.2. *As* $r \uparrow \infty$,

$$(\zeta^1(r), \ldots, \zeta^K(r)) \Rightarrow (\zeta^1, \ldots, \zeta^K) \tag{6.10}$$

in $D[0, \infty)^K$.

Before providing a proof for Theorem 6.2, we present the following two corollaries.

For $r = 1, 2 \ldots$ and $1 \leq k \leq K$, observe that

$$\begin{aligned} W_n^k(r) &= \max\{S_n^k(r) - S_i^k(r) : i = 0, 1 \ldots, n\} \\ &= S_n^k(r) - \min\{S_i^k(r) : i = 0, 1 \ldots, n\}, \qquad n = 0, 1 \ldots \end{aligned} \tag{6.11}$$

For $r = 1, 2 \ldots$, we now define the stochastic processes $\mu^k(r) \equiv \{\mu_t^k(r), t \geq 0\}, 1 \leq k \leq K$, with sample paths in $D[0, \infty)$ by

$$\mu_t^k(r) = \frac{W_{[rt]}^k(r)}{\sqrt{r}}, \quad 1 \leq k \leq K, \quad t \geq 0. \tag{6.12}$$

We also define the stochastic processes $\eta^k \equiv \{\eta_t^k, t \geq 0\}, 1 \leq k \leq K$, by

$$\mu_t^k = g(\zeta^k)_t, \quad 1 \leq k \leq K, \quad t \geq 0. \tag{6.13}$$

In Corollary 6.1 we show that the vector process associated with (6.12), converges weakly to a K–dimensional diffusion process (6.13) with drift c and covariance (6.9). This limiting diffusion stays in the non–negative orthant of the K–dimensional space and exhibits normal reflections at the boundaries.

Corollary 6.1. *As* $r \uparrow \infty$,

$$(\mu^1(r), \ldots, \mu^K(r)) \Rightarrow (\mu^1, \ldots, \mu^K) \tag{6.14}$$

in $D[0, \infty)^K$.

Proof. From (6.11) and (6.12), we conclude for each $r = 1, 2, \ldots$, that

$$\mu^k(r) = g(\zeta^k(r)), \quad 1 \le k \le K.$$

and the result follows by the continuous mapping theorem and Theorem 6.2. ∎

For $r = 1, 2 \ldots$, define the stochastic processes $\kappa(r) \equiv \{\kappa_t(r), t \ge 0\}$ with sample paths in $D[0, \infty)$ by

$$\kappa_t(r) = \frac{T_{[rt]}(r)}{\sqrt{r}}, \quad t \ge 0. \tag{6.15}$$

Also define the stochastic process $\kappa \equiv \{\kappa_t, t \ge 0\}$ with sample paths in $D[0, \infty)$ by

$$\kappa_t = \max_{1 \le k \le K} \mu_t^k, \quad t \ge 0. \tag{6.16}$$

In Corollary 7.6.2 we show that the stochastic process (6.15) generated by the end-to-end delays, converges weakly to the process (6.16), which is the maximum of K correlated Wiener processes with drift, in the non–negative orthant and normal reflection at the boundaries.

Corollary 6.2. *As $r \uparrow \infty$,*

$$\kappa(r) \Rightarrow \kappa \tag{6.17}$$

in $D[0, \infty)$.

Proof. From (6.2), (6.15) and (6.16) we conclude for each $r = 1, 2 \ldots$, that

$$\kappa_t(r) = \max_{1 \le k \le K} \{\mu_t^k(r) + \frac{a_{[rt]}^k v_{[rt]}^k}{\sqrt{r}}\}, \quad t \ge 0.$$

Equation (6.17) now follows from Corollary 6.1 by the continuous mapping theorem and the converging together theorem ∎

We now proceed with the proof of Theorem 6.2.

Proof. Note that we write (6.7) as

$$\zeta_t^k(r) = \frac{S_{[rt]}^k(r) - (p_k(r)v_k(r) - u(r))[rt]}{\sqrt{r}} + (p_k(r)v_k(r) - u(r))\frac{[rt]}{\sqrt{r}},$$

$$1 \le k \le K, \quad t \ge 0 \quad (6.18)$$

As a result of assumption **(Ik)**, we have that

$$\lim_{r\uparrow\infty}(p_k(r)v_k(r) - u(r))\frac{[rt]}{\sqrt{r}} = -c_k t. \tag{6.19}$$

By a multi–dimensional version of Prohorov's theorem, it follows that as $r \uparrow \infty$,

$$\left(\frac{S^1_{[rt]}(r) - (p_1(r)v_1(r) - u(r))[rt]}{\sqrt{r}}, \ldots, \frac{S^K_{[rt]}(r) - (p_K(r)v_K(r) - u(r))[rt]}{\sqrt{r}}\right)$$

$$\Rightarrow (\sum_{k=1}^{K} Q_{1k}\xi^1_t, \ldots, \sum_{j=k}^{K} Q_{Kk}\xi^K_t) \tag{6.20}$$

in $D[0,\infty)^K$, and it now remains for us to identify the components of the matrix Q. This can be done by observing that

$$(QQ^T)_{ij}t = R_{ij}t$$

$$= \lim_{r\uparrow\infty} \mathbb{E}\left(\frac{S^i_{[rt]}(r) - (p_i(r)v_i(r) - u(r))[rt]}{\sqrt{r}}\right)\left(\frac{S^j_{[rt]}(r) - (p_j(r)v_j(r) - u(r))[rt]}{\sqrt{r}}\right). \tag{6.21}$$

A straightforward computation of the right hand side in (6.21) leads to the conclusion that

$$R_{ij} = \begin{cases} \sigma_0^2 + p_i\sigma_i^2 + p_i(1-p_i)v_i^2, & \text{if } i = j \\ \sigma_0^2 - p_i p_j v_i v_j, & \text{if } i \ne j \end{cases}$$

which proves the theorem. ∎

6.3 Symmetric queues with Poisson arrivals

Consider the case when each queue has identical parameters so that $v = v_k, 1 \le k \le K$, $\sigma = \sigma_k, 1 \le k \le K$ and $c = c_k, 1 \le k \le K$. Further assume that $p_k = \frac{1}{K}, 1 \le k \le K$. Due to the structure of the matrix R, the assumption

$$\sigma_0^2 = \frac{v^2}{K^2} \tag{6.23}$$

leads to a cancellation of the cross–correlation terms in R, so that

$$R = \begin{pmatrix} \sigma_0^2 + \frac{\sigma^2}{K} + \frac{K-1}{K^2}v^2 & 0 & \dots & 0 \\ 0 & \sigma_0^2 + \frac{\sigma^2}{K} + \frac{K-1}{K^2}v^2 & \dots & v^2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_0^2 + \frac{\sigma^2}{K} + \frac{K-1}{K^2}v^2 \end{pmatrix}. \qquad (6.24)$$

One of the most common inter–arrival distributions that satisfies (6.23) is the exponential, since in this case

$$\sigma_0^2 = \lim_{\lambda \uparrow K\mu} \frac{1}{\lambda^2} = \frac{1}{K^2\mu^2}$$

where as usual, $v = \frac{1}{\mu}$. Making these substitutions in (6.24), we obtain

$$R = \begin{pmatrix} \frac{1}{K}(\sigma^2 + \frac{1}{\mu^2}) & 0 & \dots & 0 \\ 0 & \frac{1}{K}(\sigma^2 + \frac{1}{\mu^2}) & \dots & v^2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{K}(\sigma^2 + \frac{1}{\mu^2}) \end{pmatrix}. \qquad (6.25)$$

Thus under the condition that the arrivals are Poisson and $p_k = \frac{1}{K}, 1 \le k \le K$, (6.8) simplifies to

$$\zeta_t^k = \sqrt{\frac{1}{K}(\sigma^2 + \frac{1}{\mu^2})}\xi_t^k - ct, \quad 1 \le k \le K, \quad t \ge 0. \qquad (6.26)$$

Note that now the stochastic processes $\zeta^k, 1 \le k \le K$, are independent, so that we have reduced the diffusion to a form from which it is easy to obtain the stationary distribution. Carrying out the calculations for the case $c > 0$ as in [28], it can be shown that the RV κ_t converges in distribution to a RV κ_∞, such that

$$\mathbb{E}\kappa_\infty = (\sigma^2 + \frac{1}{\mu^2})\frac{H_K}{2Kc} \qquad (6.27)$$

where H_K as usual is the Harmonic series. In general, the n^{th} moment is given by

$$\mathbb{E}\kappa_\infty^n = n! \left[(\sigma^2 + \frac{1}{\mu^2})\frac{1}{2Kc}\right]^n \sum_{k=1}^{K} \binom{K}{k} \frac{(-1)^{k+1}}{k^n}. \qquad (6.28)$$

Let us denote the average end–to–end delay of a K–dimensional resequencing system with Poisson arrivals by $\overline{T}_K(\lambda)$ and its n^{th} moment by $\overline{T}_K^{(n)}(\lambda)$. Then (6.27)–(6.28) suggest the following formulae for the heavy traffic limit of the response times.

$$\lim_{\lambda \uparrow K\mu} (K\mu - \lambda)\overline{T}_K(\lambda) = (\sigma^2 + \frac{1}{\mu^2})\frac{KH_K\mu^2}{2} \tag{6.29}$$

and

$$\lim_{\lambda \uparrow K\mu} (K\mu - \lambda)^n \overline{T}_K^{(n)}(\lambda) = n! \left[(\sigma^2 + \frac{1}{\mu^2})\frac{K\mu^2}{2}\right]^n \sum_{k=1}^{K} \binom{K}{k} \frac{(-1)^{k+1}}{k^n}. \tag{6.30}$$

Note that in this case, in contrast to the case of disordering by $GI/GI/K$ queues, the resequencing delay grows logarithmically with K.

It is interesting to contrast the behavior of parallel queues operating under the fork–join and resequencing constraints. From [28] and (6.27) we come to the conclusion that the average response time in heavy traffic for these queues under both the synchronization constraints varies logarithmically with the number of queues K. However in light traffic, the fork–join synchronization still leads to logarithmic increase of the average response time with K (see [29]), while the resequencing synchronization leads to a constant light traffic limit for the average response time, i.e. $\overline{T}(0) = \frac{1}{\mu}$. From this we come to the conclusion that while the fork–join constraint leads to an equal degradation of the average response time inboth light and heavy traffic, the resequencing constraint becomes important in the calculation of the average response time only when the system is heavily loaded and the number of parallel queues K is large.

Before closing this section, we would like to give another example of a system in which the resequencing constraint leads to a significant degradation of the average response time in heavytraffic. The disordering system is composed of a single server queue with feedback in which the customers may be routed back to the end of the queue with probability q, or they may enter the resequencing box with probability p after receiving service. We have been unable to carry out a heavy traffic analysis of this system, however an exact analysis was carried out by Horlatt and Mailles [12] for the special case when the inter-arrival and service times are exponential with rate λ and μ respectively. They showed that the average end-to-end delay $\overline{T}(\rho)$ as a function of $\rho = \frac{\lambda}{p\mu}$ satisfies

$$\overline{T}(\rho) = \frac{p}{\mu}\sum_{k=1}^{\infty}\frac{kq^{k-1}}{[1-\rho(1-q^k)][1-\rho(1-q^{k-1})]}. \tag{6.31}$$

From (6.31) it is possible to obtain the following heavy traffic limit

$$\lim_{\rho\uparrow 1}(1-\rho)^2\overline{T}(\rho) = \sum_{k=1}^{\infty}\frac{kq^{k-1}}{(1+q^k+q^{k-1}-q^{2k-1})}. \tag{6.32}$$

Hence as a result of the resequencing constraint, the average response time grows at rate $\frac{1}{(1-\rho)^2}$ as $\rho \uparrow 1$,rather than at rate $\frac{1}{(1-\rho)}$.

REFERENCES

[1] F. Baccelli, E Gelenbe and B. Plateau,"An end-to-end approach to the resequencing problem," *JACM*, Vol. 31, No. 3, pp. 474-485 (1984).

[2] F. Baccelli and A.M. Makowski,"Queueing systems with synchronization constraints," *Proceeding of the IEEE*, (1989).

[3] F. Baccelli and A.M. Makowski,"Queueing systems with synchronization constraints," *Proceeding of the IEEE*, (1989).

[4] P. Billingsley, *"Convergence of probability measures,"* J. Wiley and Sons, New York, (1968).

[5] E.G. Coffman Jr. and M.I. Reiman,"Diffusion approximations for computer/ communication systems," *Mathematical Computer Performance and Reliability*, G. Iazeolla, P.J. Courtois and A. Hordijk (editors), North Holland (1984).

[6] L.Gün and A. Jean-Marie,"Parallel queues with resequencing," Manuscript, University of Maryland, College Park, (1988).

[7] J.M. Harrison,"The diffusion approximation for tandem queues in heavy traffic," *Adv. Appl. Prob.*, No. 10, pp. 886-905 (1978).

[8] J.M. Harrison and M.I. Reiman,"On the distribution of multidimensional reflected Brownian motion," *SIAM J. Appl. Math.*, Vol. 41, No. 2, pp. 345-361 (1981).

[9] J.M. Harrison and M.I. Reiman, "Reflected Brownian motion on an orthant," *Ann. Prob.*, Vol. 9, No. 2, pp. 302-308 (1981).

[10] J.M. Harrison, *Brownian motion and stochastic flow systems*, J. Wiley and Sons, New York (1985).

[11] J.M. Harrison and R.J. Williams,"Brownian models of open queueing networks with homogeneous customer population," *Stochastics*, Vol. 22, pp. 77-115 (1987).

[12] E. Horlatt and D. Mailles,"Etude du resequencement dans un reseau de files d'attente" *Technical Report No. 125, Universitie P. et M. Curie*, (1986).

[13] D.L. Iglehart and W. Whitt, "Multiple channel queues in heavy traffic. I," *Adv. Appl. Prob.*, No. 2, pp. 150-177 (1970).

[14] D.L. Iglehart and W. Whitt, "Multiple channel queues in heavy traffic. II: Sequences, networks and batches," *Adv. Appl. Prob.*, No. 2, pp.355-369 (1970).

[15] J.F.C. Kingman, "Queue disciplines in heavy traffic," *Math. of Oper. Res.*, Vol. 7, No. 2, pp. 262-271, (1982).

[16] J. Kollerstrom,"Heavy traffic theory for queues with several servers. I," *J. Appl. Prob.*, Vol. 11, pp. 544-552 (1974).

[17] E. Kyprianou, "The virtual waiting time of the $GI/G/1$ queue in heavy traffic," *Adv. Appl. Prob.*, No. 3, pp. 249-268 (1971).

[18] A.J. Lemoine, "Networks of queues-A survey of weak convergence results," *Manag. Sci.*, Vol. 24, No. 11, pp. 1175-1193 (1978).

[19] D.V. Lindley,"The theory of queues with a single server," *J. Appl. Prob.*, Vol. 10, pp. 109-121 (1973).

[20] T. Lindvall,"Weak convergence of probability measures and random functions in the function space $D[0,\infty)$," *J. Appl. Prob.*, Vol. 10, pp. 109-121 (1973).

[21] K.R. Parthasarthy, *Probability measures on metric spaces*, Academic Press, New York (1967).

[22] Y. Prohorov,"Convergence of random processes and limit theorems of probability theory," *Theor. Probability Appl.*, Vol. 1, pp. 157-214 (1956).

[23] M.I. Reiman, "Open queueing networks in heavy traffic," *Maths. of Oper. Res.*, Vol. 9, No. 3, pp. 441-458 (1984).

[24] S. Varma, "Some problems in queueing systems with resequencing," MS Thesis, University of Maryland; also available as SRC Technical Report No. TR-87-192 (1987).

[25] S. Varma and A.M. Makowski," Resequencing systems, Part I: Structural Properties," In preparation.

[26] S. Varma and A.M. Makowski," Resequencing systems, Part III: Interpolation approximations," In preparation.

[27] S. Varma,"Performance evaluation of the time–stamp ordering algorithm in a distributed database," Manuscript (1990).

[28] S. Varma and A.M. Makowski, "Heavy traffic limits for fork–join queues," In preparation, (1990)

[29] S. Varma and A.M. Makowski, "Interpolation approximations for fork–join queues," In preparation, (1990).

[30] W.Whitt,"Weak convergence theorems for queues in heavy traffic," Ph.D. Thesis, Cornell University (1968).

[31] W. Whitt,"Weak convergence of probability measures on the function space $C[0,\infty)$." *Ann. Math. Stat.*, Vol. 41, No. 3, pp. 939-944 (1970).

[32] W. Whitt,"Heavy traffic limit theorems for queues: A survey," Lecture Notes in Economics and Mathematical Systems, No. 98, Springer-Verlag, Berlin, pp. 307-350 (1974).

[33] W. Whitt,"Some useful functions for functional limit theorems," *Maths. of Oper. Res*, Vol. 5, No. 1, pp. 67-85 (1980). No. 98, Springer-Verlag, Berlin, pp. 307-350 (1974).

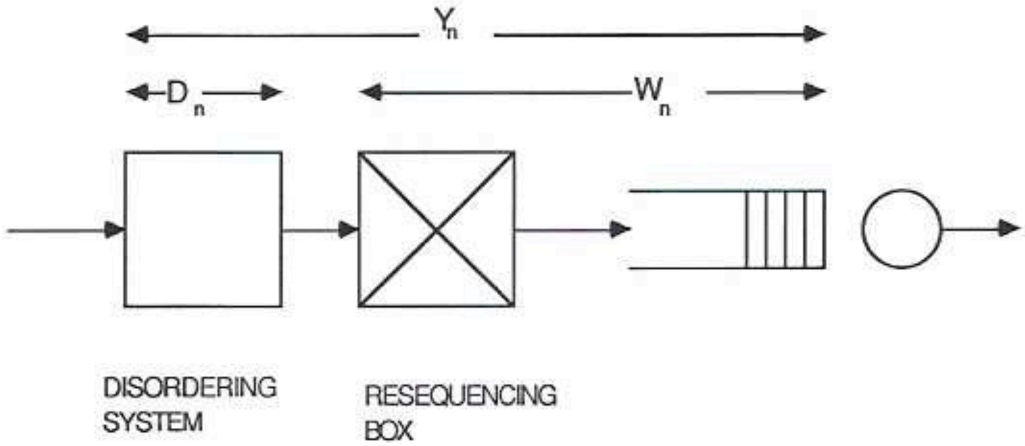

Fig. 1. A generic resequencing model

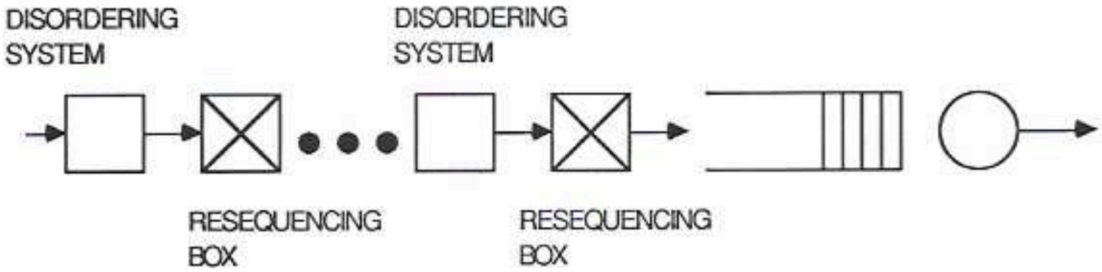

Fig. 2. A multi-stage BGP model

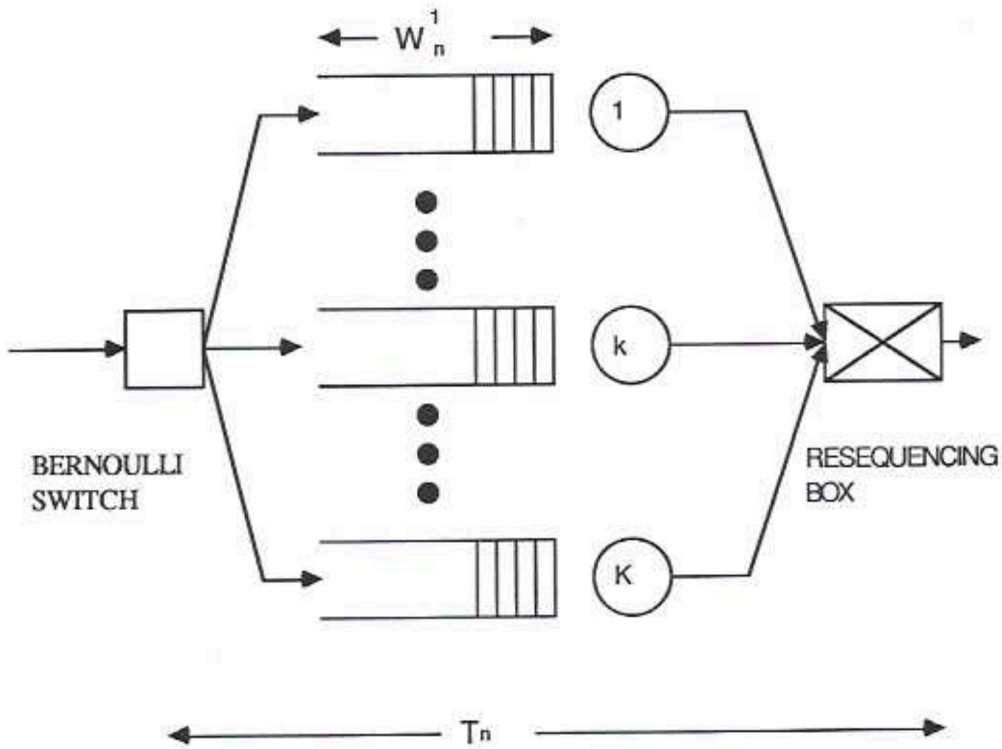

Fig. 3. Parallel queues with resequencing