On the Large Deviation of Resequencing Queue Size: 2-M/M/1 Case

Ye Xia
Department of Computer and Information
Science and Engineering
University of Florida
Gainesville, FL 32611-6120
Email: yx1@cise.ufl.edu

David Tse
Department of Electrical Engineering
and Computer Science
University of California, Berkeley
Berkeley, CA 94720–1770
Email: dtse@eecs.berkeley.edu

Abstract— **Protocols such as TCP require packets to be accepted, i.e., delivered to the receiving application, in the order they are transmitted at the sender. Packets that arrive at the receiving host may be mis-ordered for reasons such as retransmission of dropped packets or multi-path routing. In order to deliver the arrived packets in sequence, the receiver's transport layer is responsible to temporarily buffer out-of-order packets and to resequence them as more packets arrive. In this paper, we analyze a model where the mis-ordering is caused by multi-path routing. Packets are generated according to a Poisson process. Then, they arrive at a disordering network modelled by two parallel M/M/1 queues, and are routed to each of the queues according to an independent Bernoulli process. A resequencing buffer follows the disordering network. In such a model, the packet resequencing delay is known. However, the size of the resequencing queue is unknown. We derive the probability for the large deviation of the queue size.**

I. Introduction

Reliable transport protocols such as TCP requires packets to be accepted, i.e., delivered to the receiving application, in the order they are transmitted at the sender. Packets that arrive at the receiving host may be mis-ordered for several reasons, for instance, retransmission of dropped packets, or multi-path routing. The transport layer at the receiver is responsible to temporarily buffer out-of-order packets and to resequence all packets, as a result, delaying some of them. In our earlier paper [12], we model packet mis-ordering by adding an IID random propagation delay to each packet and derive simple expressions for the required buffer size and the resequencing delay. We demonstrate that these two quantities can be significant and show that the resequencing problem becomes worse as the link speed increases. In this paper, we analyze a model with correlated delays where the mis-ordering is caused by multi-path routing. Packets are generated according to a Poisson process. Then, they arrive at a disordering network modelled by two parallel M/M/1 queues, and are routed to each of the queues according to an independent Bernoulli process. A resequencing buffer follows the disordering network. In such a model, the packet resequencing delay is known. However, the size of the resequencing queue is unknown. We derive the probability for the large deviation of the queue size.

This paper is organized as follows. In Section II, we describe the resequencing model and give the main theorem of the paper. We also discuss the relation of this study with previous studies. Sections III, IV and V constitute the bulk of the paper, which is a proof for the main theorem. We show some implications of the theorem in the concluding section, VI.

II. The Model and the Main Result

The detailed network and resequencing model is shown in Figure 1. Sequentially-numbered customers (or packets) arrive at the disordering network (DN) according to a Poisson process with rate λ. Each customer either enters queue 1 with probability p, or enters queue 2 with probability $1-p$, independent of other customers. Then, the arrival processes to the queues in the DN are independent Poisson processes with rate λ_i, $i \in \{1,2\}$, where

$$\lambda_1 = p\lambda \qquad \lambda_2 = (1-p)\lambda$$

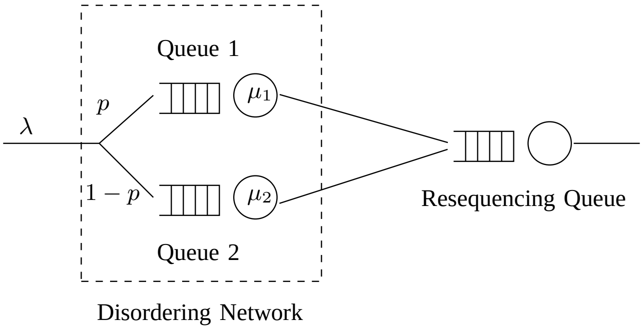

Fig. 1. Network and resequencing model

The service times for the customers at queue i are IID exponentially distributed with mean $1/\mu_i$, $i = 1,2$. Hence, we have two M/M/1 queues in the DN. Due to the multi-path routing, customers may be disordered after the DN. They are resequenced at the resequencing queue (RSQ) that follows the DN. Customers immediately leave the RSQ after they are properly resequenced. That is, customer j leaves the RSQ as soon as all customers $i < j$ have arrived at the RSQ. Note that the server of the RSQ is assumed to have infinite processing capacity. We are interested in computing the stationary queue size of the RSQ. Let q^r be the stationary size of the RSQ. The main result of this paper is the following theorem. Without the loss of generality, let us assume $\mu_1 - \lambda_1 \leq \mu_2 - \lambda_2$. Then,

Theorem 1:

$$\begin{aligned} &\lim_{n\to\infty} \tfrac{1}{n} \log P\{q^r(t) \geq n\} \\ &= \max\{\log \tfrac{\lambda_2}{\lambda_2+\mu_1-\lambda_1}, \log \tfrac{4\lambda_1\mu_1}{(\lambda_1+\mu_1+\mu_2-\lambda_2)^2}\} \end{aligned} \tag{1}$$

The studies that deal with packet mis-ordering due to multi-path routing (also including parallel processing or load balancing, etc.) typically analyze an open queueing network, of which the model in Figure 1 is a special case. In some models, a FIFO queue follows the resequencing buffer. The DN is also modelled as a queueing system, whose type typically distinguishes different studies. For instance, the DN is an M/M/∞ queue in [9], an M/GI/∞ queue in [6], a GI/GI/∞ queue in [1], an M/M/2 queue in [11], an M/M/K queue in [13], an M/H_2/K queue in [3], an M/M/2 queue with a threshold-type server assignment policy in [7], two parallel M/M/1 queues with additional fixed propagation delays in [5], and K parallel M/GI/1 queues in [8]. A survey is given in [2]. Most of these studies are concerned mostly with finding the distribution and/or mean of the resequencing delay or end-to-end delay. Several also give results about the number of packets in the resequencing queue. Among the previous studies reviewed here, the most relevant one is [8], where the DN consists of K parallel M/GI/1 queues. In [8], Jean-Marie and Gun derive the distribution of the resequencing delay. In contrast, our results are (i) for the resequencing queue size, (ii) of the large-deviation type, and (iii) for the 2-M/M/1-queue case.

In the remaining part of the paper, we will prove Theroem 1. The proof is a formalization of the following line of thoughts. Suppose the oldest customer in the DN is C_* and is being serviced at queue 1 in the DN. We wish to find out the probability that the RSQ has at least n customers. The customers in the RSQ must all arrived at the DN after C_*, and all went through queue 2 in the DN during the time C_* spent in queue 1, which is (roughly) an exponential random variable, independent of the queue 2 process. Therefore, the probability that the RSQ has at least n customers is the same as the probability that at least n customers arrive at queue 2, an M/M/1 queue, and at least n of those customers depart the queue during an exponential random time T that is independent of the queue 2 process. There is also the symmetric case where the oldest customer is in queue 2 and all customers in the RSQ come from queue 1. In Section III, we set up the two different cases and write the quantities to be computed. In Section IV, we compute the key quantity, $P\{M(T) \geq n\}$, where the function $M(t)$ is the number of those customers who arrived at the M/M/1 queue on the interval $[0, t]$ and who departed by time t, and T is an exponential random variable independent of the M/M/1 queue. In Section V, we combine results of the previous two sections and give the proof for Theroem 1.

III. The Setup

At time t, let $V(t)$ be the event {the DN is empty at time t}. If $\bar{V}(t)$, let $C_*(t)$ be the oldest customer in the DN, let $W_*(t)$ be the time $C_*(t)$ has spent in the DN, and let $I_*(t)$ be the queue in the DN which $C_*(t)$ goes through. For $n \geq 0$, let

$$\begin{aligned} E(t, s, n) \quad = \quad &\{\text{at least } n \text{ customers arrived at the DN} \\ &\text{on the interval } (t - s, t]\text{, out of which} \\ &\text{at least } n \text{ have left the DN by } t\} \end{aligned}$$

Let the size of the resequencing queue (RSQ) at time t be $q^r(t)$, and let $q_i(t)$ be the size of queue i at time t, where $i = 1$ or 2. Then, for $n > 0$,

$$P\{q^r(t) \geq n\} = P\{\bar{V}(t) \text{ and } E(t, W_*(t), n)\} \tag{2}$$

We explain the above equality in words. When the RSQ size is greater than or equal to n, where $n > 0$, it must be waiting for some customer still in the DN. In particular, the next packet gap the RSQ is trying to fill is $C_*(t)$. The customers in the RSQ are exactly those who arrived at the DN later than $C_*(t)$, but who have left the DN by time t. We are interested in computing $\lim_{t\to\infty} P\{q^r(t) \geq n\}$. Alternatively, let us assume all relevant processes are stationary.

Let us extend the definition of $W_*(t)$, $W_*(t) = 0$ if $V(t)$. Then, when $n = 0$,

$$P\{q^r(t) \geq n\} = 1$$

$$\begin{aligned} &P\{\bar{V}(t) \text{ and } E(t, W_*(t), n)\} \\ = \quad &P\{E(t, W_*(t), n)|\bar{V}(t)\}P\{\bar{V}(t)\} = P\{\bar{V}(t)\} \end{aligned}$$

$$\begin{aligned} &P\{V(t) \text{ and } E(t, W_*(t), n)\} \\ = \quad &P\{E(t, W_*(t), n)|V(t)\}P\{V(t)\} = P\{V(t)\} \end{aligned}$$

Hence, for $n = 0$,

$$P\{q^r(t) \geq n\} = P\{E(t, W_*(t), n)\} \tag{3}$$

For $n > 0$, (3) is still true because

$$\begin{aligned} &P\{V(t) \text{ and } E(t, W_*(t), n)\} \\ = \quad &P\{E(t, 0, n)|V(t)\}P\{V(t)\} = 0 \end{aligned}$$

Note that, because customers are served on first-come-first-serve basis in each of the queues, the oldest customers in the non-empty DN must be in service at one of the queues. If queue i is not empty, $i \in \{0, 1\}$, let $W_i(t)$ be the duration for which the customer in service at queue i has stayed in the queue. If queue i is empty, let $W_i(t) = 0$. By using a simple reversibility argument, $W_i(t)$ has the same distribution as the waiting time in queue i (not including the service time) by an arbitrary customer. This distribution and the density are (page 213 in [10]), for $x \geq 0$,

$$F_{W_i}(x) = P\{W_i(t) \leq x\} = 1 - \rho_i e^{-(\mu_i - \lambda_i)x} \tag{4}$$

$$f_{W_i}(x) = (1 - \rho_i)\delta(x) + \lambda_i(1 - \rho_i)e^{-(\mu_i - \lambda_i)x} \tag{5}$$

where $\rho_i = \lambda_i/\mu_i$, and $\delta(x)$ is the Dirac delta function, representing the point probability mass at $x = 0$. We will occasionally omit the dependency on t for brevity.

Let $\hat{M}_i(t,s)$ be the number of those customers who arrived at queue i on the interval $(t-s,t]$ and who departed by time t. Note that for $n>0$,

$$\begin{aligned} & P\{\hat{M}_1(t,W_*(t)) \geq n \mid W_1(t)=W_2(t)=0\} \\ = \; & P\{\hat{M}_1(t,0) \geq n \mid W_1(t)=W_2(t)=0\} = 0 \end{aligned}$$

Also,

$$P\{W_1(t)=W_2(t) \neq 0\} = 0$$

Therefore,

$$\begin{aligned} & P\{\hat{M}_1(t,W_*(t)) \geq n \mid W_1(t)=W_2(t)\} \\ & \cdot P\{W_1(t)=W_2(t)\} = 0 \end{aligned}$$

Then, for $n>0$,

$$\begin{aligned} & P\{q^r(t) \geq n\} \\ = \; & P\{E(t,W_*(t),n)\} \\ = \; & P\{\hat{M}_2(t,W_*(t)) \geq n \mid W_1(t) > W_2(t)\} \\ & \cdot P\{W_1(t) > W_2(t)\} \\ & + P\{\hat{M}_1(t,W_*(t)) \geq n \mid W_2(t) > W_1(t)\} \\ & \cdot P\{W_2(t) > W_1(t)\} \end{aligned}$$

This can be explained as follows. If $W_1(t) > W_2(t)$, then the oldest customer, $C_*(t)$, in the DN must be in service at queue 1. Hence, $W_1(t) = W_*(t)$. All customers who came to the DN after $C_*(t)$ and who have left the DN by time t must have been routed to the RSQ via queue 2.

For $n>0$,

$$\begin{aligned} & P\{\hat{M}_2(t,W_*(t)) \geq n \mid W_1(t) > W_2(t)\} \\ = \; & \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n \mid W_1(t)=s, W_1(t) > W_2(t)\} \\ & \cdot f_{W_1|W_1>W_2}(s) ds \\ = \; & \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n \mid W_1(t)=s, W_2(t) < s\} \\ & \cdot f_{W_1|W_1>W_2}(s) ds \\ = \; & \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n \mid W_2(t) < s\} \\ & \cdot f_{W_1|W_1>W_2}(s) ds \end{aligned} \tag{6}$$

In the above, $f_{W_1|W_1>W_2}(s)$ denote the conditional density of $W_1(t)$ given $\{W_1(t) > W_2(t)\}$. In the last step, we used the fact that the two queue processes are independent. Note that, in the integral, the (conditional) probability mass at $s=0$ does not contribute to the probability on the left hand side.

We will compute the conditional density by starting with the joint probability. For $x \geq 0$,

$$\begin{aligned} & P\{W_1 > x, W_1 > W_2\} \\ = \; & \rho_1 e^{-(\mu_1-\lambda_1)x} - \rho_1\rho_2 \frac{\mu_1-\lambda_1}{\mu_1-\lambda_1+\mu_2-\lambda_2} \\ & \cdot e^{-(\mu_1-\lambda_1+\mu_2-\lambda_2)x} \end{aligned} \tag{7}$$

From (7), we have

$$\begin{aligned} P\{W_1 > W_2\} &= P\{W_1 > 0, W_1 > W_2\} \\ &= \rho_1 - \rho_1\rho_2 \frac{\mu_1-\lambda_1}{\mu_1-\lambda_1+\mu_2-\lambda_2} \end{aligned} \tag{8}$$

From (7) and (8), we get the conditional density for $x \geq 0$.

$$\begin{aligned} & f_{W_1|W_1>W_2}(x) \\ = \; & K_1 e^{-(\mu_1-\lambda_1)x} - K_2 e^{-(\mu_1-\lambda_1+\mu_2-\lambda_2)x} \end{aligned} \tag{9}$$

where K_1 and K_2 are constants, given by,

$$K_1 = \frac{\mu_1-\lambda_1}{1-\rho_2 \frac{\mu_1-\lambda_1}{\mu_1-\lambda_1+\mu_2-\lambda_2}}$$

$$K_2 = \frac{\rho_2(\mu_1-\lambda_1)}{1-\rho_2 \frac{\mu_1-\lambda_1}{\mu_1-\lambda_1+\mu_2-\lambda_2}}$$

Note that the second term in (9) decays much faster than the first term. If we ignore it, the conditional probability density decays exponentially.

Next, we will bound (6) from above and below.

$$\begin{aligned} & \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n \mid W_2(t) < s\} f_{W_1|W_1>W_2}(s) ds \\ = \; & \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n, W_2(t) < s\} \frac{f_{W_1|W_1>W_2}(s)}{P\{W_2(t) < s\}} ds \\ \leq \; & \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n\} \frac{f_{W_1|W_1>W_2}(s)}{P\{W_2(t) = 0\}} ds \\ \leq \; & \frac{1}{1-\rho_2} \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n\} f_{W_1|W_1>W_2}(s) ds \end{aligned} \tag{10}$$

For a lower bound,

$$\begin{aligned} & \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n \mid W_2(t) < s\} f_{W_1|W_1>W_2}(s) ds \\ = \; & \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n, W_2(t) < s\} \frac{f_{W_1|W_1>W_2}(s)}{P\{W_2(t) < s\}} ds \\ \geq \; & \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n, W_2(t) = 0\} f_{W_1|W_1>W_2}(s) ds \\ = \; & \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n, q_2(t) = 0\} \\ & \cdot f_{W_1|W_1>W_2}(s) ds \end{aligned} \tag{11}$$

In the next section, we will prepare to compute the upper and lower bound.

IV. Computation of $P\{M(T) \geq n\}$

In this section, we consider a stationary M/M/1 queue whose arrival rate is λ_1 and whose departure rate is μ_1. We assume $\lambda_1 < \mu_1$ so that the queue is stable. Let T be an exponential random variable independent of the queue process with mean $1/(\mu_2-\lambda_2)$, where $\lambda_2 < \mu_2$. Let $M(t)$ be the number of those customers who arrived on the interval $[0,t]$ and who departed by time t. We wish to compute $P\{M(T) \geq n\}$ for large n. The main result of this section is Theorem 2.

Theorem 2:

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\log P\{M(T)\geq n\}\\ &= \begin{cases}\log\frac{\lambda_1}{\lambda_1+\mu_2-\lambda_2} & \text{if } \mu_1-\lambda_1\geq\mu_2-\lambda_2\\ \log\frac{4\lambda_1\mu_1}{(\lambda_1+\mu_1+\mu_2-\lambda_2)^2} & \text{if } \mu_1-\lambda_1<\mu_2-\lambda_2\end{cases}\end{aligned} \tag{12}$$

Lemma 3:

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\log P\{M(T)\geq n, q(T)=0\}\\ &\geq \begin{cases}\log\frac{\lambda_1}{\lambda_1+\mu_2-\lambda_2} & \text{if } \mu_1-\lambda_1\geq\mu_2-\lambda_2\\ \log\frac{4\lambda_1\mu_1}{(\lambda_1+\mu_1+\mu_2-\lambda_2)^2} & \text{if } \mu_1-\lambda_1<\mu_2-\lambda_2\end{cases}\end{aligned} \tag{13}$$

Proof: This is an intermediate step in the proof of lower bound for Theorem 2. ■

In the next two subsections, we will prove the Theorem 2. We will frequently use the following fact. For $a>0$ and integer $k\geq 0$,

Fact 4:

$$\int_0^\infty \frac{e^{-at}t^k}{k!}dt = (\frac{1}{a})^{k+1} \tag{14}$$

A. Case of $\mu_1-\lambda_1\geq\mu_2-\lambda_2$

1) The Upper Bound:

$$\begin{aligned}& P\{M(T)\geq n\}\\ \leq\ & P\{\text{the number of customer arrivals on the}\\ & \text{interval } [0,T] \text{ is at least } n\} \qquad (15)\\ =\ & \sum_{k=n}^\infty\int_0^\infty\frac{e^{-\lambda_1 t}(\lambda_1 t)^k}{k!}(\mu_2-\lambda_2)e^{-(\mu_2-\lambda_2)t}dt\\ =\ & \sum_{k=n}^\infty(\mu_2-\lambda_2)\int_0^\infty\frac{e^{-(\lambda_1+\mu_2-\lambda_2)t}(\lambda_1 t)^k}{k!}dt\\ =\ & \sum_{k=n}^\infty\frac{\mu_2-\lambda_2}{\lambda_1+\mu_2-\lambda_2}(\frac{\lambda_1}{\lambda_1+\mu_2-\lambda_2})^k\\ =\ & \frac{\mu_2-\lambda_2}{\lambda_1+\mu_2-\lambda_2}\frac{(\frac{\lambda_1}{\lambda_1+\mu_2-\lambda_2})^n}{1-\frac{\lambda_1}{\lambda_1+\mu_2-\lambda_2}}\\ =\ & (\frac{\lambda_1}{\lambda_1+\mu_2-\lambda_2})^n\end{aligned}$$

2) The Lower Bound: Suppose, as n gets large,

$$\begin{aligned}&\int_0^\infty\frac{e^{-\lambda_1 t}(\lambda_1 t)^n}{n!}(\mu_2-\lambda_2)e^{-(\mu_2-\lambda_2)t}dt\\ \approx\ & \max_{t\geq 0}\frac{e^{-\lambda_1 t}(\lambda_1 t)^n}{n!}(\mu_2-\lambda_2)e^{-(\mu_2-\lambda_2)t}\end{aligned}$$

in some sense. It can be shown easily that the integrand is maximized at,

$$t_o = n/(\lambda_1+\mu_2-\lambda_2) \tag{16}$$

This information will be useful in the proof for the lower bound.

Let $q(t)$ be the queue size at time t. Let $D(t)$ be the number of departures on the interval $[0,t]$.

$$\begin{aligned}& P\{M(t)=k\}\\ =\ & \sum_{m=0}^\infty P\{M(t)=k|q(0)=m\}P\{q(0)=m\}\\ \geq\ & P\{M(t)=k|q(0)=0\}P\{q(0)=0\}\\ =\ & (1-\rho_1)P\{D(t)=k|q(0)=0\}\\ \geq\ & (1-\rho_1)P\{D(t)=k,q(t)=0|q(0)=0\}\end{aligned} \tag{17}$$

From [4] (page 199),

$$\begin{aligned}& P\{D(t)=k,q(t)=0|q(0)=0\}\\ =\ & \sum_{i=0}^\infty\frac{(1+i)\rho_1^k}{k!(k+i+1)!}(\mu_1 t)^{2k+i}e^{-(\lambda_1+\mu_1)t}\\ =\ & \frac{(\lambda_1 t)^k e^{-\lambda_1 t}}{k!}\sum_{i=0}^\infty\frac{1+i}{(k+i+1)!}(\mu_1 t)^{k+i}e^{-\mu_1 t}\\ \geq\ & \frac{1}{k+1}\frac{(\lambda_1 t)^k e^{-\lambda_1 t}}{k!}\sum_{i=0}^\infty\frac{1}{(k+i)!}(\mu_1 t)^{k+i}e^{-\mu_1 t}\\ =\ & \frac{1}{k+1}\frac{(\lambda_1 t)^k e^{-\lambda_1 t}}{k!}P\{Y_{(\mu_1 t)}\geq k\}\end{aligned} \tag{18}$$

where $Y_{(\mu_1 t)}$ is a Poisson random variable with mean $\mu_1 t$. Now, with the definition of t_o as in (16),

$$\begin{aligned}& P\{M(T)\geq n\}\\ \geq\ & P\{M(T)\geq n, T\geq t_o\}\\ \geq\ & P\{M(t_o)\geq n, T\geq t_o\}\\ =\ & P\{M(t_o)\geq n\}P\{T\geq t_o\} \qquad (19)\\ =\ & \sum_{k=n}^\infty P\{M(t_o)=k\}P\{T\geq t_o\} \qquad (20)\end{aligned}$$

The equality in (19) is because of independence between the queue process and the random variable T. Then, by (20), (17) and (18),

$$\begin{aligned}& P\{M(T)\geq n\}\\ \geq\ & (1-\rho_1)\sum_{k=n}^\infty\frac{1}{k+1}\frac{(\lambda_1 t_o)^k e^{-\lambda_1 t_o}}{k!}\\ & \cdot P\{Y_{(\mu_1 t_o)}\geq k\}e^{-(\mu_2-\lambda_2)t_o}\\ \geq\ & (1-\rho_1)\frac{1}{n+1}\frac{(\lambda_1 t_o)^n e^{-\lambda_1 t_o}}{n!}\\ & \cdot P\{Y_{(\mu_1 t_o)}\geq n\}e^{-(\mu_2-\lambda_2)t_o}\end{aligned} \tag{21}$$

We will show $P\{Y_{(\mu_1 t_o)}\geq n\}$ is greater than a constant as n tends to infinity. By the definition of t_o and by the assumption $\mu_1-\lambda_1\geq\mu_2-\lambda_2$,

$$\mu_1 t_o = \frac{\mu_1}{\lambda_1+\mu_2-\lambda_2}n\geq n$$

Let $n_o=\lfloor\mu_1 t_o\rfloor$. Then, $n_o\geq n$. Let $X_1, X_2, ..., X_{n_o}$ be IID.

Poisson random variables with mean 1. Then,

$$\begin{aligned} P\{Y_{(\mu_1 t_o)} \geq n\} &\geq P\{\frac{X_1+X_2+...+X_{n_o}}{n_o} \geq \frac{n}{n_o}\} \\ &\geq P\{\frac{X_1+X_2+...+X_{n_o}}{n_o} \geq 1\} \\ &= P\{\frac{X_1+X_2+...+X_{n_o}-n_o}{\sqrt{n_o}\sqrt{n_o}} \geq 0\} \end{aligned}$$

By the central limit theorem,

$$\begin{aligned} &\lim_{n_o\to\infty} P\{\frac{X_1+X_2+...+X_{n_o}-n_o}{\sqrt{n_o}\sqrt{n_o}} \geq 0\} \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{2} \end{aligned}$$

Therefore, for any $\epsilon > 0$, there exists some integer $N > 0$ such that for all $n > N$,

$$P\{Y_{(\mu_1 t_o)} \geq n\} \geq \frac{1}{2} - \epsilon \tag{22}$$

Continuing from (21), for all $n > N$,

$$\begin{aligned} &P\{M(T) \geq n\} \\ \geq\ &(1-\rho_1)(\frac{1}{2}-\epsilon)\frac{1}{n+1}\frac{(\lambda_1 t_o)^n e^{-\lambda_1 t_o}}{n!} e^{-(\mu_2-\lambda_2)t_o} \end{aligned}$$

By the Stirling's approximation,

$$n! = \sqrt{2\pi n} n^n e^{-n}(1+O(1/n))$$

For n large enough,

$$n! \leq 2\sqrt{2\pi n} n^n e^{-n}$$

Therefore, for large enough n,

$$\begin{aligned} &P\{M(T) \geq n\} \\ \geq\ &\frac{1}{4}(1-\rho_1)(1-2\epsilon)\frac{1}{n+1}(\frac{\lambda_1}{\lambda_1+\mu_2-\lambda_2})^n \\ &\cdot \frac{n^n \exp(-\frac{\lambda_1}{\lambda_1+\mu_2-\lambda_2}n)}{\sqrt{2\pi n} n^n e^{-n}} \exp(-\frac{\mu_2-\lambda_2}{\lambda_1+\mu_2-\lambda_2}n) \\ =\ &\frac{(1-\rho_1)(1-2\epsilon)}{4\sqrt{2\pi n}(n+1)}(\frac{\lambda_1}{\lambda_1+\mu_2-\lambda_2})^n \end{aligned} \tag{23}$$

B. Case of $\mu_1 - \lambda_1 < \mu_2 - \lambda_2$

1) The Lower Bound: By (17) and (18),

$$\begin{aligned} &P\{M(T) \geq n\} \\ \geq\ &\int_0^\infty \sum_{k=n}^\infty (1-\rho_1) P\{D(t)=k, q(t)=0 | q(0)=0\} \\ &\cdot (\mu_2-\lambda_2) e^{-(\mu_2-\lambda_2)t} dt \\ \geq\ &(1-\rho_1)(\mu_2-\lambda_2)\frac{1}{n+1} \\ &\cdot \int_0^\infty \frac{(\lambda_1 t)^n e^{-\lambda_1 t}}{n!} P\{Y_{(\mu_1 t)} = n\} e^{-(\mu_2-\lambda_2)t} dt \\ =\ &(1-\rho_1)(\mu_2-\lambda_2)\frac{1}{n+1} \\ &\cdot \int_0^\infty \frac{(\lambda_1 t)^n e^{-\lambda_1 t}}{n!}\frac{(\mu_1 t)^n e^{-\mu_1 t}}{n!} e^{-(\mu_2-\lambda_2)t} dt \\ =\ &(1-\rho_1)(\mu_2-\lambda_2)\frac{1}{n+1}\frac{(2n)!}{n!n!}(\lambda_1\mu_1)^n \\ &\cdot \int_0^\infty \frac{t^{2n} e^{-(\lambda_1+\mu_1+\mu_2-\lambda_2)t}}{(2n)!} dt \\ =\ &(1-\rho_1)\frac{\mu_2-\lambda_2}{\lambda_1+\mu_1+\mu_2-\lambda_2}\frac{1}{n+1}\frac{(2n)!}{n!n!} \\ &\cdot (\lambda_1\mu_1)^n \frac{1}{(\lambda_1+\mu_1+\mu_2-\lambda_2)^{2n}} \end{aligned}$$

By the Stirling's approximation, for large enough n,

$$\frac{(2n)!}{n!n!} = \frac{\sqrt{4\pi n}(2n)^{2n} e^{-2n}(1+O(1/n))}{(\sqrt{2\pi n}(n)^n e^{-n}(1+O(1/n)))^2} \geq \frac{C_1}{\sqrt{n}} 4^n$$

for some constant $C_1 > 0$. Therefore,

$$\begin{aligned} &P\{M(T) \geq n\} \\ \geq\ &C_1(1-\rho_1)\frac{\mu_2-\lambda_2}{\lambda_1+\mu_1+\mu_2-\lambda_2} \\ &\cdot \frac{1}{\sqrt{n}(n+1)}(\frac{4\lambda_1\mu_1}{(\lambda_1+\mu_1+\mu_2-\lambda_2)^2})^n \end{aligned} \tag{24}$$

2) The Upper Bound: The computation for the upper bound in the previous case does not work here. To see the reason, consider the integral in the lower bound calculation. Suppose, as n becomes large,

$$\begin{aligned} &\int_0^\infty \frac{(\lambda_1 t)^n e^{-\lambda_1 t}}{n!}\frac{(\mu_1 t)^n e^{-\mu_1 t}}{n!} e^{-(\mu_2-\lambda_2)t} dt \\ \approx\ &\max_{t\geq 0} \frac{(\lambda_1 t)^n e^{-\lambda_1 t}}{n!}\frac{(\mu_1 t)^n e^{-\mu_1 t}}{n!} e^{-(\mu_2-\lambda_2)t} \end{aligned}$$

It can be shown easily the above maximum is achieved at

$$t_o = \frac{2n}{\lambda_1+\mu_1+\mu_2-\lambda_2} \tag{25}$$

Note that when $\mu_1 - \lambda_1 < \mu_2 - \lambda_2$,

$$\mu_1 t_0 = \frac{2\mu_1 n}{\lambda_1+\mu_1+\mu_2-\lambda_2} < n$$

Therefore, $\{Y_{(\mu_1 t_o)} \geq n\}$ is a large deviation type of event instead of an event with constant probability, as n becomes large. It is not tight enough to bound $P\{M(t) \geq n\}$ from above by only looking at the arrival processes, as was done in (15).

$$\begin{aligned} &P\{M(t) \geq n\} \\ \leq\ &P\{\text{at least } n \text{ customers arrived on the interval} \\ &[0,t]\text{, and at least } n \text{ customers are served on} \\ &\text{the same interval}\} \\ \leq\ &\sum_{k=n}^\infty \frac{e^{-\lambda_1 t}(\lambda_1 t)^k}{k!} P\{\sum_{i=1}^n X_i \leq t\} \end{aligned} \tag{26}$$

where $\{X_1, X_2, ..., X_n\}$ are IID service times. The sum $\sum_{i=1}^n X_i$ has the Gamma distribution with density,

$$f(t) = \frac{\mu_1 e^{-\mu_1 t}(\mu_1 t)^{n-1}}{(n-1)!}$$

Hence,

$$\begin{aligned}
& P\{M(T) \geq n\} \\
\leq \;& \sum_{k=n}^{\infty} \int_0^{\infty} \frac{e^{-\lambda_1 t}(\lambda_1 t)^k}{k!} \\
& \int_0^t \frac{\mu_1 e^{-\mu_1 \tau}(\mu_1 \tau)^{n-1}}{(n-1)!} d\tau (\mu_2 - \lambda_2) e^{-(\mu_2-\lambda_2)t} dt \\
= \;& \mu_1(\mu_2 - \lambda_2) \sum_{k=n}^{\infty} \int_0^{\infty} \int_{\tau}^{\infty} \frac{e^{-(\lambda_1+\mu_2-\lambda_2)t}(\lambda_1 t)^k}{k!} dt \\
& \cdot \frac{e^{-\mu_1 \tau}(\mu_1 \tau)^{n-1}}{(n-1)!} d\tau
\end{aligned}$$

Let $t = \tau + u$. The above becomes,

$$\begin{aligned}
& P\{M(T) \geq n\} \\
\leq \;& \mu_1(\mu_2 - \lambda_2) \sum_{k=n}^{\infty} \int_0^{\infty} \int_0^{\infty} \\
& \frac{e^{-(\lambda_1+\mu_2-\lambda_2)(\tau+u)}(\lambda_1(\tau+u))^k}{k!} du \frac{e^{\mu_1 \tau}(\mu_1 \tau)^{n-1}}{(n-1)!} d\tau \\
= \;& \mu_1(\mu_2 - \lambda_2) \sum_{k=n}^{\infty} \lambda_1^k \int_0^{\infty} \int_0^{\infty} \\
& \frac{e^{-(\lambda_1+\mu_2-\lambda_2)u} \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} u^i \tau^{k-i}}{k!} du \\
& \frac{e^{-(\lambda_1+\mu_1+\mu_2-\lambda_2)\tau}(\mu_1 \tau)^{n-1}}{(n-1)!} d\tau \\
= \;& \mu_1(\mu_2 - \lambda_2) \sum_{k=n}^{\infty} \lambda_1^k \sum_{i=0}^{k} \frac{1}{(k-i)!} \\
& \int_0^{\infty} \int_0^{\infty} \frac{e^{-(\lambda_1+\mu_2-\lambda_2)u} u^i}{i!} du \\
& \frac{e^{-(\lambda_1+\mu_1+\mu_2-\lambda_2)\tau} \tau^{k-i} (\mu_1 \tau)^{n-1}}{(n-1)!} d\tau \\
= \;& (\mu_2 - \lambda_2)\mu_1^n \sum_{k=n}^{\infty} \lambda_1^k \\
& \sum_{i=0}^{k} \frac{1}{(k-i)!} \frac{1}{(\lambda_1+\mu_2-\lambda_2)^{i+1}} \\
& \int_0^{\infty} \frac{e^{-(\lambda_1+\mu_1+\mu_2-\lambda_2)\tau} \tau^{n-1+k-i}}{(n-1)!} d\tau \\
= \;& (\mu_2 - \lambda_2)\mu_1^n \sum_{k=n}^{\infty} \lambda_1^k \sum_{i=0}^{k} \frac{(n-1+k-i)!}{(k-i)!(n-1)!} \\
& \frac{1}{(\lambda_1+\mu_2-\lambda_2)^{i+1}} \frac{1}{(\lambda_1+\mu_1+\mu_2-\lambda_2)^{n+k-i}} \\
= \;& \frac{\mu_2 - \lambda_2}{\lambda_1+\mu_2-\lambda_2} \left(\frac{\mu_1}{\lambda_1+\mu_1+\mu_2-\lambda_2}\right)^n \\
& \sum_{k=n}^{\infty} \left(\frac{\lambda_1}{\lambda_1+\mu_1+\mu_2-\lambda_2}\right)^k \\
& \cdot \sum_{i=0}^{k} \frac{(n-1+k-i)!}{(k-i)!(n-1)!} \left(\frac{\lambda_1+\mu_1+\mu_2-\lambda_2}{\lambda_1+\mu_2-\lambda_2}\right)^i
\end{aligned} \tag{27}$$

For $i = 0, 1, ..., k$, define

$$a(k,i) = \frac{(n-1+k-i)!}{2^k (k-i)!(n-1)!}$$

Let

$$\beta = \frac{\lambda_1+\mu_1+\mu_2-\lambda_2}{\lambda_1+\mu_2-\lambda_2}$$

Note that for $\mu_1 - \lambda_1 < \mu_2 - \lambda_2$, $\beta < 2$.

$$\frac{a(k+1,i)}{a(k,i)} = \frac{n+k-i}{2(k+1-i)} = \frac{1 + \frac{n-1}{k+1-i}}{2}$$

Then, for each fixed $i \in \{0, 1, ..., k\}$,

$$\frac{a(k+1,i)}{a(k,i)} \begin{cases} \geq 1 & \text{if } k \leq n+i-2 \\ < 1 & \text{if } k > n+i-2 \end{cases}$$

Therefore, $a(k,i)$ is maximized at $k = n+i-1$ for each i [1]. Then,

$$a(n+i-1, i) = \frac{(2n-2)!}{(n-1)!(n-1)!2^{n+i-1}}$$

Then, the sum in (27) index by i becomes,

$$\begin{aligned}
& \sum_{i=0}^{k} \frac{(n-1+k-i)!}{(k-i)!(n-1)!} \left(\frac{\lambda_1+\mu_1+\mu_2-\lambda_2}{\lambda_1+\mu_2-\lambda_2}\right)^i \\
= \;& 2^k \sum_{i=0}^{k} a(k,i)\beta^i \\
\leq \;& 2^{k-n} \sum_{i=0}^{k} \frac{2(2n-2)!}{(n-1)!(n-1)!} \left(\frac{\beta}{2}\right)^i \\
\leq \;& 2^{k-n} \frac{2(2n-2)!}{(n-1)!(n-1)!} \sum_{i=0}^{\infty} \left(\frac{\beta}{2}\right)^i \\
= \;& 2^{k-n} \frac{2(2n-2)!}{(n-1)!(n-1)!} \frac{2}{2-\beta}
\end{aligned}$$

The infinite sum above is finite because $\beta/2 < 1$. Going back to (27), we get,

$$\begin{aligned}
& P\{M(T) \geq n\} \\
\leq \;& \frac{4(\mu_2 - \lambda_2)}{(2-\beta)(\lambda_1+\mu_2-\lambda_2)} \frac{(2n-2)!}{2^n (n-1)!(n-1)!} \\
& \left(\frac{\mu_1}{\lambda_1+\mu_1+\mu_2-\lambda_2}\right)^n \sum_{k=n}^{\infty} \left(\frac{2\lambda_1}{\lambda_1+\mu_1+\mu_2-\lambda_2}\right)^k \\
= \;& \frac{4(\mu_2 - \lambda_2)}{\mu_2 - \lambda_2 - (\mu_1 - \lambda_1)} \frac{(2n-2)!}{2^n (n-1)!(n-1)!} \\
& \left(\frac{\mu_1}{\lambda_1+\mu_1+\mu_2-\lambda_2}\right)^n \\
& \cdot \left(\frac{2\lambda_1}{\lambda_1+\mu_1+\mu_2-\lambda_2}\right)^n / \left(1 - \frac{2\lambda_1}{\lambda_1+\mu_1+\mu_2-\lambda_2}\right)
\end{aligned}$$

The above sum is finite because, for $\mu_1 - \lambda_1 < \mu_2 - \lambda_2$ and $\lambda_1 < \mu_1$, we have

$$\frac{2\lambda_1}{\lambda_1+\mu_1+\mu_2-\lambda_2} < 1 \tag{28}$$

[1] We assume that n is large enough when necessary. In this case, $n \geq 1$.

Next, Stirling's approximation yields,

$$\begin{aligned}\frac{(2n)!}{n!n!} &= \frac{\sqrt{4\pi n}(2n)^{2n}e^{-2n}(1+O(1/n))}{(\sqrt{2\pi n}(n)^n e^{-n}(1+O(1/n)))^2} \\ &\le \frac{C_2}{\sqrt{n}}4^n \end{aligned} \tag{29}$$

for some constant $C_2 > 0$. Hence, for some other constant $C_4 > 0$, we have,

$$P\{M(T) \ge n\} \quad \le \quad \frac{C_4}{\sqrt{n}}(\frac{4\lambda_1\mu_1}{(\lambda_1+\mu_1+\mu_2-\lambda_2)^2})^n$$

V. Proof of Theorem 1

We will combine the results of the previous two sections and prove the main theorem. We wish to show that, without the loss of generality, when $\mu_1 - \lambda_1 \le \mu_2 - \lambda_2$,

$$\begin{aligned}&\lim_{n\to\infty} \tfrac{1}{n}\log P\{q^r(t) \ge n\} \\ &= \max\{\log \tfrac{\lambda_2}{\lambda_2+\mu_1-\lambda_1}, \log \tfrac{4\lambda_1\mu_1}{(\lambda_1+\mu_1+\mu_2-\lambda_2)^2}\}\end{aligned} \tag{30}$$

Proof: Start with (11).

$$\begin{aligned}&\int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \ge n \mid W_2(t) < s\} f_{W_1|W_1>W_2}(s)ds \\ \ge\ & \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \ge n, q_2(t) = 0\} f_{W_1|W_1>W_2}(s)ds \\ =\ & \int_{0^+}^{\infty} P\{M_2(s) \ge n, q_2(s) = 0\} f_{W_1|W_1>W_2}(s)ds \\ =\ & K_1 \int_{0}^{\infty} P\{M_2(s) \ge n, q_2(s) = 0\} e^{-(\mu_1-\lambda_1)s}ds \\ & -K_2 \int_{0}^{\infty} P\{M_2(s) \ge n, q_2(s) = 0\} \\ & e^{-(\mu_1-\lambda_1+\mu_2-\lambda_2)s}ds \end{aligned} \tag{31}$$

where $K_1 > 0$ and $K_2 > 0$ are some constants. In the above, we used the fact the conditional density of W_1 given $\{W_1 > W_2\}$ takes the form as in (9). By Lemma 13 with suitable substitution of variables, since

$$\mu_2 - \lambda_2 < \mu_1 - \lambda_1 + \mu_2 - \lambda_2$$

the second term in (31) gives

$$\begin{aligned}&\lim_{n\to\infty} \frac{1}{n}\log \int_0^{\infty} P\{M_2(s) \ge n, q_2(s) = 0\} \\ &\cdot e^{-(\mu_1-\lambda_1+\mu_2-\lambda_2)s}ds \\ \ge\ & \log \frac{4\lambda_2\mu_2}{(\lambda_2+\mu_2+\mu_1-\lambda_1+\mu_2-\lambda_2)^2} \\ \ge\ & \log \frac{4\lambda_2\mu_2}{(2\mu_2+\mu_1-\lambda_1)^2}\end{aligned} \tag{32}$$

The first term in (31) gives

$$\begin{aligned}&\lim_{n\to\infty} \frac{1}{n}\log \int_0^{\infty} P\{M_2(s) \ge n, q_2(s) = 0\} \\ &\cdot e^{-(\mu_1-\lambda_1)s}ds \\ \ge\ & \begin{cases} \log \frac{\lambda_2}{\lambda_2+\mu_1-\lambda_1} & \text{if } \mu_2-\lambda_2 \ge \mu_1-\lambda_1 \\ \log \frac{4\lambda_2\mu_2}{(\lambda_2+\mu_2+\mu_1-\lambda_1)^2} & \text{if } \mu_2-\lambda_2 < \mu_1-\lambda_1 \end{cases}\end{aligned} \tag{33}$$

Now,

$$\begin{aligned}&\frac{\lambda_2}{\lambda_2+\mu_1-\lambda_1} \\ =\ & \frac{4\lambda_2\mu_2}{4\lambda_2\mu_2+4\mu_1\mu_2-4\lambda_1\mu_2}\end{aligned}$$

$$\begin{aligned}&\frac{4\lambda_2\mu_2}{(2\mu_2+\mu_1-\lambda_1)^2} \\ =\ & \frac{4\lambda_2\mu_2}{4\mu_2^2+4\mu_1\mu_2-4\lambda_1\mu_2+(\mu_1-\lambda_1)^2}\end{aligned}$$

Hence,

$$\frac{\lambda_2}{\lambda_2+\mu_1-\lambda_1} \ge \frac{4\lambda_2\mu_2}{(2\mu_2+\mu_1-\lambda_1)^2}$$

Also, because $\lambda_2 < \mu_2$,

$$\frac{4\lambda_2\mu_2}{(\lambda_2+\mu_2+\mu_1-\lambda_1)^2} \ge \frac{4\lambda_2\mu_2}{(2\mu_2+\mu_1-\lambda_1)^2}$$

Therefore, we can ignore the contribution from (32) when considering the lower bound. Then, (33) gives the correct lower bound. Using similarly argument and by Theorem 2, we get

$$\begin{aligned}&\lim_{n\to\infty} \tfrac{1}{n}\log \textstyle\int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \ge n\} f_{W_1|W_1>W_2}(s)ds \\ &= \begin{cases} \log \frac{\lambda_2}{\lambda_2+\mu_1-\lambda_1} & \text{if } \mu_2-\lambda_2 \ge \mu_1-\lambda_1 \\ \log \frac{4\lambda_2\mu_2}{(\lambda_2+\mu_2+\mu_1-\lambda_1)^2} & \text{if } \mu_2-\lambda_2 < \mu_1-\lambda_1 \end{cases}\end{aligned} \tag{34}$$

Note that

$$\begin{aligned}& P\{\hat{M}_2(t,s) \ge n\} \\ =\ & P\{\hat{M}_2(t,s) \ge n \mid W_2(t) < s\} P\{\ W_2(t) < s\} \\ & + P\{\hat{M}_2(t,s) \ge n \mid W_2(t) \ge s\} P\{\ W_2(t) \ge s\} \\ \ge\ & P\{\hat{M}_2(t,s) \ge n \mid W_2(t) < s\}\end{aligned}$$

Hence, (34) is an upper bound of

$$\int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \ge n \mid W_2(t) < s\} f_{W_1|W_1>W_2}(s)ds$$

Since the upper and lower bound agree with each other, by (6), we get

$$\begin{aligned}&\lim_{n\to\infty} \tfrac{1}{n}\log P\{\hat{M}_2(t, W_*(t)) \ge n \mid W_1(t) > W_2(t)\} \\ &= \begin{cases} \log \frac{\lambda_2}{\lambda_2+\mu_1-\lambda_1} & \text{if } \mu_2-\lambda_2 \ge \mu_1-\lambda_1 \\ \log \frac{4\lambda_2\mu_2}{(\lambda_2+\mu_2+\mu_1-\lambda_1)^2} & \text{if } \mu_2-\lambda_2 < \mu_1-\lambda_1 \end{cases}\end{aligned} \tag{35}$$

To determine $P\{q^r(t) \ge n\}$ for large n, we also need to compute $P\{\hat{M}_1(t, W_*(t)) \ge n \mid W_2(t) > W_1(t)\}$ (See (6)). By symmetry,

$$\begin{aligned}&\lim_{n\to\infty} \tfrac{1}{n}\log P\{\hat{M}_1(t, W_*(t)) \ge n \mid W_2(t) > W_1(t)\} \\ &= \begin{cases} \log \frac{\lambda_1}{\lambda_1+\mu_2-\lambda_2} & \text{if } \mu_1-\lambda_1 \ge \mu_2-\lambda_2 \\ \log \frac{4\lambda_1\mu_1}{(\lambda_1+\mu_1+\mu_2-\lambda_2)^2} & \text{if } \mu_1-\lambda_1 < \mu_2-\lambda_2 \end{cases}\end{aligned} \tag{36}$$

When $\mu_1 - \lambda_1 \le \mu_2 - \lambda_2$, combining (35), (36) and (6), we get (30). ■

VI. Conclusion

Some remarks about Theorem 1 are as follows. When, $\lambda_1 = \lambda_2$ and $\mu_1 = \mu_2$,

$$\lim_{n\to\infty} \frac{1}{n} \log P\{q^r(t) \geq n\} = \log \rho_1$$

When $\mu_1 - \lambda_1 = \mu_2 - \lambda_2$,

$$\lim_{n\to\infty} \frac{1}{n} \log P\{q^r(t) \geq n\} = \max\{\log \rho_1, \log \rho_2\}$$

Like all GI/GI/1 queues, the resequencing queue size depends on the arrival and departure rates through a dimensionless parameter. This implies that the resequencing queue size does not change with the link speed of the network, assuming the traffic characteristics are not altered by the technology change. This is in contrast with the models from our previous paper [12], where the improvement of network speed worsens the packet resequencing problem in terms of both the queue size and the delay. In the current model, there can be many ways to produce the large resequencing queue size, which, in general, depends on parameters for both queues in the DN. An interesting observation is that it can be large even when the queue sizes in the DN are both small. This occurs when the two disordering queues are “mismatched”. For example, suppose $\mu_i = 2\lambda_i$ for $i = 1$ and 2. Hence, $\rho_1 = \rho_2 = 1/2$, and for $i = 1$ and 2,

$$\lim_{n\to\infty} \frac{1}{n} \log P\{q_i(t) \geq n\} = \log \frac{1}{2}$$

Suppose $\lambda_2 = 10\lambda_1$. Then,

$$\lim_{n\to\infty} \frac{1}{n} \log P\{q^r(t) \geq n\} = \log \frac{10}{11}$$

In Figure 2, we show the simulation results for $P\{q^r = n\}$ and compare them with the analytical results in Theorem 1. In Figure 2 (a), the parameters are chosen so that

$$\frac{\lambda_2}{\lambda_2 + \mu_1 - \lambda_1} = \frac{10}{11} = 0.9091$$

$$\frac{4\lambda_1\mu_1}{(\lambda_1 + \mu_1 + \mu_2 - \lambda_2)^2} = \frac{8}{169} = 0.0473$$

In Figure 2 (b), the parameters are chosen so that

$$\frac{\lambda_2}{\lambda_2 + \mu_1 - \lambda_1} = \frac{1}{21} = 0.0476$$

$$\frac{4\lambda_1\mu_1}{(\lambda_1 + \mu_1 + \mu_2 - \lambda_2)^2} = \frac{800}{1681} = 0.4759$$

Loosely speaking, Theorem 1 says, for $\mu_1 - \lambda_1 \leq \mu_2 - \lambda_2$ and for large n,

$$P\{q^r(t) \geq n\} = e^{-\delta n + o(n)} \tag{37}$$

where $o(n)$ is a function that grows more slowly than n, i.e., $o(n)/n \to 0$ as n tends to infinity. The large deviation analysis of this paper is able to give an expression for the parameter δ,

$$\delta = -\max\{\log \frac{\lambda_2}{\lambda_2 + \mu_1 - \lambda_1}, \log \frac{4\lambda_1\mu_1}{(\lambda_1 + \mu_1 + \mu_2 - \lambda_2)^2}\}$$

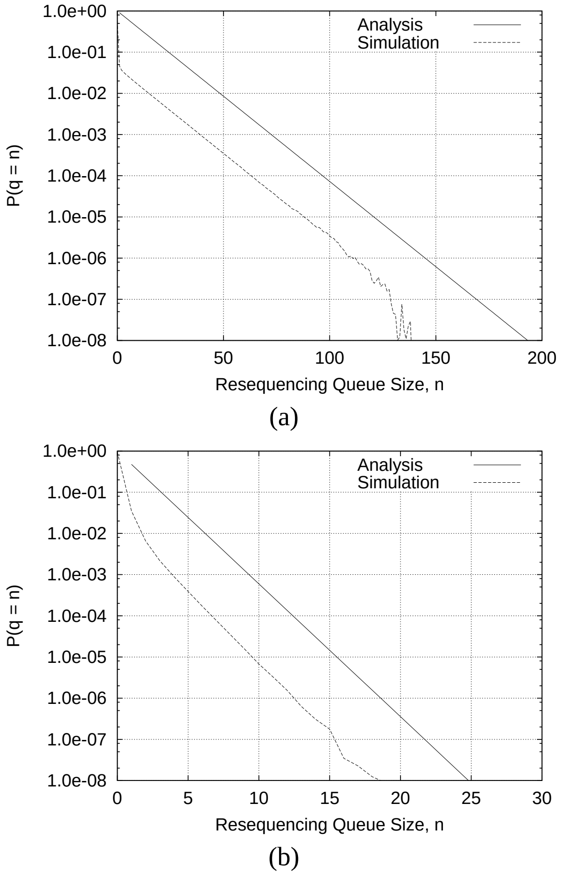

Fig. 2. $P\{q^r = n\}$: Simulation results. (a) $\lambda_1 = 1$, $\mu_1 = 2$, $\lambda_2 = 10$, $\mu_2 = 20$; (b) $\lambda_1 = 10$, $\mu_1 = 20$, $\lambda_2 = 1$, $\mu_2 = 12$

but cannot capture the nature of $o(n)$. In each plot of Figure 2, the gap between the two curves shows the “imprecision” of the large deviation result. That is, it shows how much the large deviation result misses the actual tail probability of the queue size.

From the modelling point of view, compared with those in [12], the model in this paper allows non-IID packet delays in the DN and it specifically models packet disordering caused by routing on different paths. The analysis is generalizable to more complex arrival and service processes for the queues in the DN, even non-IID arrival processes. One weakness of this model is that it does not allow situations that yield heavy-tailed distributions for the RSQ.

References

[1] F. Baccelli, E. Gelenbe, and B. Plateau. An end-to-end approach to the resequencing problem. *Journal of the Association for Computing Machinery, Vol. 31, No. 3,* pages 474–485, July 1984.

[2] F. Baccelli and A. M. Makowski. Queueing models for systems with synchronization constraints. *Proceedings of the IEEE, Vol. 77, No. 1,* pages 138–161, January 1989.

[3] S. Chowdhury. An analysis of virtual circuits with parallel links. *IEEE Transactions on Communications, Vol. 39, No. 8,* pages 1184–1188, August 1991.

[4] J. W. Cohen. *The Single Server Queue*. North-Holland Publishing Company, 1st edition, 1969.

[5] N. Gogate and S. S. Panwar. On a resequencing model for high speed networks. In *Proceedings of INFOCOM '94*, pages 40–47, Toronto, Canada, June 1994.

[6] G. Harrus and B. Plateau. Queueing analysis of a reordering issue. *IEEE Transaction on Software Engineering, Vol. SE-8, No. 2*, pages 113–123, March 1982.

[7] I. Iliadis and L. Y.-C. Lien. Resequencing delay for a queueing system with two heterogeneous servers under a threshold-type scheduling. *IEEE Transactions on Communications, Vol. 36, No. 6*, pages 692–702, June 1988.

[8] A. Jean-Marie and L. Gün. Parallel queues with resequencing. *Journal of the Association for Computing Machinery, Vol. 40, No. 5*, pages 1188–1208, November 1993.

[9] F. Kamoun, L. Kleinrock, and R. Muntz. Queueing analysis of the reordering issue in a distributed database concurrency control mechanism. In *Proceedings of the 2nd International Conference on Distributed Computing Systems*, pages 13–23, Versailles, France, April 1981.

[10] Leonard Kleinrock. *Queue Systems, Volume I: Theory*. Jonh Wiley & Sons, 1975.

[11] Y.-C. Lien. Evaluation of the resequence delay in a Poisson queueing system with two heterogeneous servers. In *Proceedings of the International Workshop on Computer Performance Evaluation*, pages 189–197, Tokyo, Japan, September 1985.

[12] Ye Xia and David Tse. Analysis on Packet Resequencing for Reliable Network Protocols. In *Proceedings of the IEEE Infocom 2003*, San Francisco, CA, April 2003.

[13] T.-S. P. Yum and T.-Y. Ngai. Resequencing of messages in communication networks. *IEEE Transactions on Communications, Vol. COM-34, No. 2*, pages 143–149, February 1986.