On the Large Deviation of Resequencing Queue Size - A Generalizable Argument involving Moment Generating Functions

I. 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

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

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 $n > 0$,

$$P\{q^r(t) \geq n\} = P\{\bar{V}(t) \text{ and } E(t, W_*(t), n)\}. \tag{1}$$

Next, we will explain equality (1). 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 $n = 0$,

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

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

$$P\{V(t) \text{ and } E(t, W_*(t), n)\}$$

$$= P\{E(t, W_*(t), n)|V(t)\}P\{V(t)\} = P\{V(t)\}.$$

Hence, for $n = 0$,

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

For $n > 0$, (2) is still true because

$$P\{V(t) \text{ and } E(t, W_*(t), n)\}$$

$$= P\{E(t, 0, n)|V(t)\}P\{V(t)\} = 0.$$

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 $W_i(t) = 0$.

We need the following result. As $x \to \infty$, for $i = 1$ or 2,

$$P\{W_i(t) > x\} \sim e^{-\delta x}, \tag{3}$$

$$f_{W_i}(x) \sim e^{-\delta x}. \tag{4}$$

Here, we assume the pdf, $f_{W_i}(x)$, exists. 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 $n > 0$,

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

$$= P\{\hat{M}_1(t, 0) \geq n \mid W_1(t) = W_2(t) = 0\} = 0.$$

Also, because packets cannot simultaneously arrive at different disordering queues,

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

Therefore,

$$P\{\hat{M}_1(t, W_*(t)) \geq n \mid W_1(t) = W_2(t)\}$$

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

Then, for $n > 0$,

$$\begin{aligned} & P\{q^r(t) \geq n\} \\ = & P\{E(t, W_*(t), n)\} \\ = & P\{\hat{M}_2(t, W_*(t)) \geq n \mid W_1(t) > W_2(t)\} \\ & \cdot P\{W_1(t) > W_2(t)\} \\ & + P\{\hat{M}_1(t, W_*(t)) \geq n \mid W_2(t) > W_1(t)\} \\ & \cdot P\{W_2(t) > W_1(t)\}. \end{aligned} \tag{5}$$

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 $n > 0$,

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

In the above, $f_{W_1|W_1>W_2}(s)$ denotes the conditional density of $W_1(t)$ given $\{W_1(t) > W_2(t)\}$. We need this to have an exponential decaying tail $e^{-\delta s}$. 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,

$$K_1 = \frac{\mu_1 - \lambda_1}{1 - \rho_2 \frac{\mu_1-\lambda_1}{\mu_1-\lambda_1+\mu_2-\lambda_2}}, \tag{10}$$

$$K_2 = \frac{\rho_2(\mu_1 - \lambda_1)}{1 - \rho_2 \frac{\mu_1-\lambda_1}{\mu_1-\lambda_1+\mu_2-\lambda_2}}. \tag{11}$$

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.

$$\begin{aligned} & \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n \mid W_1(t) = s, W_2(t) < s\} f_{W_1|W_1>W_2}(s) ds \\ &= \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n, W_2(t) < s \mid W_1(t) = s\} \frac{f_{W_1|W_1>W_2}(s)}{P\{W_2(t) < s \mid W_1(t) = s\}} ds \\ &\leq \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n \mid W_1(t) = s\} \frac{f_{W_1|W_1>W_2}(s)}{P\{W_2(t) = 0 \mid W_1(t) = s\}} ds \\ &\leq A_1 \int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n \mid W_1(t) = s\} f_{W_1|W_1>W_2}(s) ds. \end{aligned} \tag{12}$$

Here, we need the result that

$$P\{W_2(t) < s \mid W_1(t) = s\} \geq \xi > 0, \tag{13}$$

or

$$P\{W_2(t) = 0 \mid W_1(t) = s\} \geq \xi > 0, \tag{14}$$

and $A_1 = \frac{1}{\xi}$.

Next,

$$\begin{aligned} & P\{\hat{M}_2(t,s) \geq n \mid W_1(t) = s\} \\ &\leq P\{\text{at least } n \text{ customers arrive at disordering queue 2 on the interval } [t-s,t], \text{ and} \\ &\quad \text{at least } n \text{ of these customers are served on the same interval} \\ &\quad \mid W_1(t) = s\} \end{aligned} \tag{15}$$

Let the customer who is in service at time t at disordering queue i be denoted by $C_i(t)$. The event $E = \{W_1(t) = s\}$ says that $C_1(t)$ arrived at queue 1 at time $t - s$ and its sojourn time is at least s. The only relevance of the event E to the queue 2 process after time $t-s$ is that an arrival occurred at time $t-s$, which was directed to queue 1. This affects the probability of when the next arrival to queue 2 occurs. Conditional on E, starting at time $t-s$, the arrival process to queue 2 is a (non-delayed) renewal process. Also, the service times of the arrivals on the interval $[t-s,t]$ are independent of E. We denote T_i as the interarrival times and S_i as the service times of customers to queue 2. Hence, we can write the following upper bound for (15).

$$\begin{aligned}
&P\{\hat{M}_2(t,s) \geq n \mid W_1(t)=s\}\\
\leq\ &P\{\text{at least } n \text{ customers arrive at disordering queue 2 on the interval } [t-s,t]\\
&\quad \text{according to the renewal process, and}\\
&\quad \text{at least } n \text{ of these customers are served on the same interval}\}\\
\leq\ &P\{\sum_{i=1}^{n} T_i \leq s\} P\{\sum_{i=1}^{n} S_i \leq s\}. \qquad (16)
\end{aligned}$$

For a lower bound of (6),

$$\begin{aligned}
&\int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n \mid W_1(t)=s, W_2(t)<s\} f_{W_1|W_1>W_2}(s) ds\\
=&\int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n, W_2(t)<s \mid W_1(t)=s\} \frac{f_{W_1|W_1>W_2}(s)}{P\{W_2(t)<s \mid W_1(t)=s\}} ds\\
\geq&\int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n, W_2(t)=0 \mid W_1(t)=s\} f_{W_1|W_1>W_2}(s) ds\\
=&\int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n, q_2(t)=0 \mid W_1(t)=s\} f_{W_1|W_1>W_2}(s) ds\\
\geq&\int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n, q_2(t)=0, q_2(t-s)=0 \mid W_1(t)=s\} f_{W_1|W_1>W_2}(s) ds\\
=&\int_{0^+}^{\infty} P\{\hat{M}_2(t,s) \geq n, q_2(t)=0 \mid q_2(t-s)=0, W_1(t)=s\}\\
&\cdot P\{q_2(t-s)=0 \mid W_1(t)=s\} f_{W_1|W_1>W_2}(s) ds. \qquad (17)
\end{aligned}$$

Let us assume (which should be proven in general cases), for all $s \geq 0$,

$$P\{q_2(t-s)=0 \mid W_1(t)=s\} \geq A_2 > 0. \qquad (18)$$

As argued before, conditional on E, starting at time $t-s$, the arrival process to queue 2 is a renewal process. A particular way for

Also, the only relevance of the event $\{W_1(t)=s\}$ to the queue 2 process after time $t-s$ is that an arrival occurred at time $t-s$, which was directed to queue 1.

conditional on $\{q_2(t-s)=0\}$, the queue process of queue 2 after time $t-s$ is independent of $\{W_1(t)=s\}$, which is completely determined by time $t-s$. Hence, the lower bound of (6) can be written as,

$$\begin{aligned}&\int_{0^+}^{\infty} P\{\hat{M}_2(t,s)\ge n \mid W_1(t)=s, W_2(t)<s\} f_{W_1|W_1>W_2}(s)ds\\ \ge &A_2\int_{0^+}^{\infty} P\{\hat{M}_2(t,s)\ge n, q_2(t)=0 \mid q_2(t-s)=0\} f_{W_1|W_1>W_2}(s)ds. \end{aligned} \tag{19}$$

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

II.

From the analysis of the M/M/1 case, we see that the following is the important quantity.

$$\int_0^{\infty}\sum_{i=n}^{\infty}\frac{e^{-\lambda_1 t}(\lambda_1 t)^i}{i!}\sum_{l=n}^{\infty}\frac{e^{-\mu_1 t}(\mu_1 t)^l}{l!}(\mu_2-\lambda_2)e^{-(\mu_2-\lambda_2)t}dt$$

The first sum is the probability that more than n customers have arrived on the interval $[0,t]$ for a Poisson process with rate λ_1. The second sum is the probability that $\sum_{j=0}^{n} S_j \le t$, where S_j's are iid exponential random variables with mean $1/\mu_1$. This motivates the consideration of the following quantity.

$$\int_0^{\infty} P\{\sum_{i=1}^{n} T_i \le t\} P\{\sum_{j=1}^{n} S_j \le t\} e^{-\delta t}dt \tag{20}$$

where $\delta=\mu_2-\lambda_2$, T_i's are iid random variables with mean $1/\lambda_1$, representing the interarrival times, and S_i's are iid random variables with mean $1/\mu_1$, representing the service times. We'd like to study the asymptotic value of (20).

For any random variable X, let us write its moment generating function by

$$M_X(\theta)=\mathbf{E}e^{\theta X}$$

It can be verified that $M_X(\theta)$ is convex, and is increasing if $X\ge 0$. The log moment generating function $\log M_X(\theta)$ is convex, and is increasing if $X \ge 0$. Let

$$I_X(a)=\sup_{\theta}(\theta a-\log M_X(\theta)) \tag{21}$$

It can be shown that $I_X(a)$ is non-negative, convex, and differentiable(?). We will use the theory of large deviations [2]. For that purpose, we make some regularity assumptions.

H1 $M_T(\theta)<\infty$ for θ in some neighborhood of 0. For $a<1/\lambda_1$, $I_T(a)$ is achieved in the interior of the neighborhood.

H2 $M_S(\theta)<\infty$ for θ in some neighborhood of 0. For $a<1/\mu_1$, $I_S(a)$ is achieved in the interior of the neighborhood.

Let us define

$$t_1 = \frac{n}{\mu_1} \tag{22}$$

$$t_2 = \frac{n}{\lambda_1} \tag{23}$$

From the Chernoff bound, when $0 < t < t_2$, we have

$$P\{\sum_{i=1}^{n} T_i \leq t\} \leq e^{-nI_T(t/n)} \tag{24}$$

When $0 < t < t_1$, we have

$$P\{\sum_{i=1}^{n} S_i \leq t\} \leq e^{-nI_S(t/n)} \tag{25}$$

On the interval $[0, t_1)$, both $\{\sum_{i=1}^{n} T_i \leq t\}$ and $\{\sum_{j=1}^{n} S_j \leq t\}$ are large deviation events, as n increases. On $[t_1, t_2)$, $\{\sum_{i=1}^{n} T_i \leq t\}$ is a large deviation event and $\{\sum_{j=1}^{n} S_j \leq t\}$ is an event of constant probability. On $[t_2, \infty]$, both are events of constant probability. Hence, for large n, the integral in (20) can be bounded from above by

$$\int_0^{t_1} C_1(n) e^{-nI_T(t/n)} e^{-nI_S(t/n)} e^{-\delta t} dt + \int_{t_1}^{t_2} C_2(n) e^{-nI_T(t/n)} e^{-\delta t} dt + \int_{t_2}^{\infty} C_3 e^{-\delta t} dt \tag{26}$$

where the functions $C_1(n)$ and $C_2(n)$ are on the order of $e^{o(n)}$, and C_3 is a constant. The third term in (26) gives,

$$C_3 e^{-\frac{\delta}{\lambda_1} n} \tag{27}$$

where we reuse the symbol C_3 for some other constant. We next work with the other two terms in (26), and will start with the second term.

A. *Second Term*

We wish to compute

$$\int_{t_1}^{t_2} C_2(n) e^{-nI_T(t/n)} e^{-\delta t} dt = \int_{t_1}^{t_2} C_2(n) e^{-nI_T(t/n) - \delta t} dt \tag{28}$$

Let us ask what the supremum of the integrand is and where it occurs if it can be achieved. By the definition of $I_T(a)$, $I_T(a)$ and $\log M_T(\theta)$ are convex conjugate pair with,

$$\log M_T(\theta) = \sup_a (\theta a - I_T(a)) \tag{29}$$

Therefore, we have

$$\log M_T(\theta) = \sup_t (\theta \frac{t}{n} - I_T(t/n)) \tag{30}$$

Hence,

$$n \log M_T(-\delta) = \sup_t(-\delta t - nI_T(t/n)) \tag{31}$$

Suppose θ_o achieves the supremum of

$$I_T(a) = \sup_\theta(\theta a - \log M_T(\theta)) \tag{32}$$

θ_o satisfies

$$a = \frac{M'_T(\theta_o)}{M_T(\theta_o)} \tag{33}$$

Let us define a function

$$h_T(\theta) = \frac{M'_T(\theta)}{M_T(\theta)} \tag{34}$$

Then,

$$h'_T(\theta) = \frac{\mathbf{E}[T^2 e^{\theta T}]\mathbf{E}[e^{\theta T}] - (\mathbf{E}[Te^{\theta T}])^2}{(\mathbf{E}[e^{\theta T}])^2} \tag{35}$$

Here, we make the assumption

H3 $h'_T(\theta)$ exists in some neighborhood of 0.

By the Cauchy-Schwarz inequality

$$\mathbf{E}[T^2 e^{\theta T}]\mathbf{E}[e^{\theta T}] \geq (\mathbf{E}[Te^{\theta T}])^2 \tag{36}$$

where the equality holds if and only if $X = 0$ with probability 1. Hence, $h_T(\theta)$ is an increasing function. We will assume it is a strictly increasing function. With this,

$$\theta_o(a) = h_T^{-1}(a) \tag{37}$$

where $h_T^{-1}(a)$ is the inverse of h_T and is also a strictly increasing function. With θ_o, we have

$$I_T(a) = \theta_o a - \log M_T(\theta_o) \tag{38}$$

$$\begin{aligned} I'_T(a) &= a\theta'_o + \theta_o - \frac{M'_T(\theta_o)}{M_T(\theta_o)}\theta'_o \\ &= a\theta'_o + \theta_o - a\theta'_o \\ &= \theta_o \\ &= h_T^{-1}(a) \end{aligned} \tag{39}$$

Then,

$$-I'_T(a) - \delta = -h_T^{-1}(a) - \delta \tag{40}$$

We have proven the following lemma.

Lemma 1:

$$I_T'(a) = -\delta \Longleftrightarrow h_T^{-1}(a) = -\delta \Longleftrightarrow a = h_T(-\delta) = \frac{M_T'(-\delta)}{M_T(-\delta)} \tag{41}$$

$$I_T'(a) \leq -\delta \Longleftrightarrow h_T^{-1}(a) \leq -\delta \Longleftrightarrow a \leq h_T(-\delta) = \frac{M_T'(-\delta)}{M_T(-\delta)} \tag{42}$$

The inequality $\leq$ can be replaced by $<$, $>$ and $\geq$ in the above.

Returning back to (31), suppose the supremum is achieved at t_*. Then, t_* satisfies

$$-\delta - I_T'(t_*/n) = -\delta - h_T^{-1}(t_*/n) = 0 \tag{43}$$

Hence,

$$t_* = nh_T(-\delta) \tag{44}$$

Lemma 2: t_* exists, is unique, and $0 < t_* < t_2 = n/\lambda_1$.

Proof: Note that for all $\delta > 0$, $M_T'(-\delta)$ and $M_T(-\delta)$ are both defined and positive. Because h_T is strictly increasing, for $\delta > 0$,

$$0 < h_T(-\delta) = \frac{M_T'(-\delta)}{M_T(-\delta)} < h_T(0) = 1/\lambda_1 \tag{45}$$

Hence, if t_* is defined as in (44), it satisfies $0 < t_* < t_2 = n/\lambda_1$. Therefore, t_* is the unique point that achieves the supremum of (31).

■

We next consider two cases, depending on whether t_* lies in $(0, t_1)$ or $[t_1, t_2)$.

B. case 1: $t_* \in (0, t_1)$

We make additional assumption.

H4 $I_T''(a)$ is continuous on $[0, \infty)$.

Because of (39), **H4** is equivalent to

H4′ $h_T'(\theta)$ is continuous in a neighborhood of 0.

Lemma 3: $t_* \in (0, t_1)$ if and only if $h_T(-\delta) < 1/\mu_1$. When $t_* \in (0, t_1)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{t_1}^{t_2} C_2(n) e^{-nI_T(t/n)} e^{-\delta t} dt = -(I_T(1/\mu_1) + \delta/\mu_1) \tag{46}$$

Let

$$\phi(t) = e^{-nI_T(t/n) - \delta t} \tag{47}$$

We will first show that

Lemma 4: When n is large enough, $\phi(t)$ is convex and decreasing on $[t_1, t_2]$.

Proof:

$$\phi'(t) = e^{-nI_T(t/n)-\delta t}(-I'_T(t/n) - \delta) \tag{48}$$

$$\phi''(t) = e^{-nI_T(t/n)-\delta t}[(-I'_T(t/n) - \delta)^2 - \frac{1}{n} I''_T(t/n)] \tag{49}$$

We already know that

$$-I'_T(t/n) - \delta \quad \begin{cases} > 0 & \text{if } t < t_* \\ = 0 & \text{if } t = t_* \\ < 0 & \text{if } t > t_* \end{cases} \tag{50}$$

Also, $-I'_T(t/n) - \delta$ strictly decreases. Hence, on $[t_1, t_2]$,

$$-I'_T(t/n) - \delta \le -I'_T(1/\mu_1) - \delta < -I'_T(t_*/n) - \delta = 0 \tag{51}$$

Hence, $(-I'_T(t/n) - \delta)^2$ is greater than some constant, $c > 0$. On the other hand, $I''_T(t/n)$ is bounded on $[t_1, t_2]$. Hence, $I''_T(t/n)/n$ tends to 0 as n increases. Therefore, for n large enough, $\phi''(t)$ is positive on $[t_1, t_2]$. ■

Proof: (of Lemma 3) By (44), $t_* < t_1$ if and only if $nh_T(-\delta) < t_1$, and if and only if $h_T(-\delta) < 1/\mu_1$.

The function $\phi(t)$ increases when $t < t_*$ and then decreases when $t > t_*$. By examining $\phi''(t)$, due to the fact that $I''_T(a) > 0$, we know that $\phi(t)$ is concave in the neighborhood of $t = t_*$. A qualitative illustration of $\phi(t)$ is shown in figure 1. In it, we draw a tangent line at the point $(t_1, \phi(t_1))$. Let the tangent line intercepts the t-axis at t_x. Then,

$$t_x - t_1 = \frac{\phi(t_1)}{\phi'(t_1)} = \frac{1}{I'_T(1/\mu_1) + \delta} \tag{52}$$

Note that $t_x - t_1$ is a constant. Hence, for n large enough, t_x falls between t_1 and t_2.

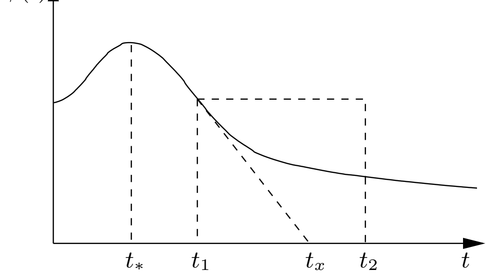

Fig. 1. $\phi(t)$. Case 1: $t_* \in (0, t_1)$.

By convexity of $\phi(t)$ on $[t_1, t_2]$, $\phi(t)$ is above the tangent line on that interval. Therefore, the integral in (28) is bounded from above by the area of the rectangle between t_1 and t_2, and is bounded below by the area of the triangle between t_1 and t_x, multiplied by $C_2(n)$. Hence,

$$\frac{1}{2}C_2(n)\phi(t_1)(t_x - t_1) \leq \int_{t_1}^{t_2} C_2(n)e^{-nI_T(t/n)-\delta t}dt \leq C_2(n)\phi(t_1)(t_2 - t_1) \tag{53}$$

$$\frac{1}{2}\frac{1}{I'_T(1/\mu_1)+\delta}C_2(n)\phi(t_1) \leq \int_{t_1}^{t_2} C_2(n)e^{-nI_T(t/n)-\delta t}dt \leq (1/\lambda_1 - 1/\mu_1)nC_2(n)\phi(t_1) \tag{54}$$

Hence,

$$\int_{t_1}^{t_2} C_2(n)e^{-nI_T(t/n)-\delta t}dt = e^{-n(I_T(1/\mu_1)+\delta/\mu_1)+o(n)} \tag{55}$$

■

Example 1:

Suppose T has exponential distribution with mean $1/\lambda_1$. Then

$$M_T(\theta) = \frac{\lambda_1}{\lambda_1 - \theta} \tag{56}$$

$$h_T(\theta) = \frac{M'_T(\theta)}{M_T(\theta)} = \frac{1}{\lambda_1 - \theta} \tag{57}$$

Hence the condition for $t_* \in (0, t_1)$ is $\frac{1}{\lambda_1+\delta} < 1/\mu_1$, which is equivalent to

$$\mu_1 - \lambda_1 < \delta \tag{58}$$

It is worth pointing out that $\mu_1 - \lambda_1 < \delta$ is not sufficient to guarantee that $t_* \in (0, t_1)$ in general situations. To see this, let us rewrite the condition for $t_* \in (0, t_1)$, $h_T(-\delta) < 1/\mu_1$, as

$$\mathbf{E}[(\frac{1}{\mu_1} - T)e^{-\delta T}] > 0 \tag{59}$$

Write $\psi(x) = (\frac{1}{\mu_1} - x)e^{-\delta x}$. Then,

$$\psi'(x) = e^{-\delta x}(-1 - \frac{\delta}{\mu_1} + \delta x) \tag{60}$$

Hence, at $x_o = \frac{1}{\mu_1} + \frac{1}{\delta}$, $\psi'(x_o) = 0$. When $x < x_o$, $\psi'(x) < 0$, and when $x > x_o$, $\psi'(x) > 0$. An illustration of $\psi(x)$ is shown in figure 2, where we choose δ to be large so that $\mu_1 - \lambda_1 < \delta$ and $1/\lambda_1 > 1/\mu_1 + 1/\delta$. In this case, if the pdf of T, denoted by $f_T(x)$, is concentrated enough around the mean $1/\lambda_1$, then $\mathbf{E}\psi(T) < 0$.

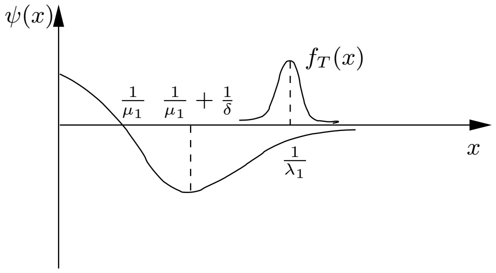

Fig. 2. A situation where $\mathbf{E}\psi(T) < 0$. $f_T(x)$ is the pdf of T.

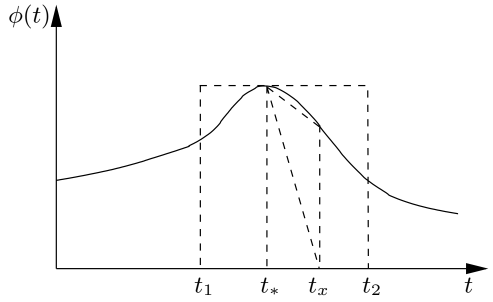

Fig. 3. $\phi(t)$. Case 2: $t_* \in [t_1, t_2)$.

C. case 2: $t_* \in [t_1, t_2)$

Lemma 5: $t_* \in [t_1, t_2)$ if and only if $h_T(-\delta) \geq 1/\mu_1$. When $t_* \in [t_1, t_2)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{t_1}^{t_2} C_2(n) e^{-nI_T(t/n)} e^{-\delta t} dt = \log M_T(-\delta) \tag{61}$$

Proof: By (44), $t_* \geq t_1$ if and only if $nh_T(-\delta) \geq t_1$, and if and only if $h_T(-\delta) \geq 1/\mu_1$.

For this case, the function $\phi(t)$ is illustrated in figure 3. We know that $\phi(t)$ is concave in a neighborhood of t_* and will become convex when t becomes large, for large enough n. Using the same argument as the proof of Lemma 4, $\phi(t)$ is convex on $[t_2, \infty)$ for large enough n. Hence, the point $t_x > t_*$ where $\phi(t)$ turns from being concave to being convex lies on (t_*, t_2).

We wish to show that $t_x - t_*$ does not become small as n increases. Because $\phi''(t_x) = 0$, we have

$$(-I_T'(t_x/n) - \delta)^2 = \frac{1}{n} I_T''(t_x/n) \tag{62}$$

Because $I_T'(a)$ is an increasing function,

$$I_T'(t_x/n) = -\delta + \sqrt{I_T''(t_x/n)}/\sqrt{n} \tag{63}$$

By **H4**, $I_T''(t/n)$ is bounded on $t \in [t_*, t_2]$. Hence, there exist constants $0 < d_1 \leq d_2$

$$d_1 \leq I_T'(t_x/n) \leq d_2 \tag{64}$$

Because $h_T(\theta)$ is an increasing function and by lemma 1,

$$h_T(-\delta + d_1/\sqrt{n}) \leq \frac{t_x}{n} \leq h_T(-\delta + d_2/\sqrt{n}) \tag{65}$$

We will expand h_T by Taylor's formula. For that purpose, we need the assumption

H5 $h_T''(\theta)$ exists and is continuous in some neighborhood of $-\delta$.

Then, for some $\xi_1 \in (-\delta, -\delta + \frac{d_1}{\sqrt{n}})$,

$$h_T(-\delta + \frac{d_1}{\sqrt{n}}) = h_T(-\delta) + h_T'(-\delta)\frac{d_1}{\sqrt{n}} + \frac{h_T''(\xi_1)}{2}(\frac{d_1}{\sqrt{n}})^2 \tag{66}$$

Hence,

$$t_x \geq nh_T(-\delta) + d_1 h_T'(-\delta)\sqrt{n} + \frac{d_1^2 h_T''(\xi_1)}{2} \tag{67}$$

Similarly,

$$t_x \leq nh_T(-\delta) + d_2 h_T'(-\delta)\sqrt{n} + \frac{d_2^2 h_T''(\xi_2)}{2} \tag{68}$$

for some $\xi_2 \in (-\delta, -\delta + \frac{d_2}{\sqrt{n}})$. By substituting $t_* = nh_T(-\delta)$, we get

$$d_1 h_T'(-\delta)\sqrt{n} + \frac{d_1^2 h_T''(\xi_1)}{2} \leq t_x - t_* \leq d_2 h_T'(-\delta)\sqrt{n} + \frac{d_2^2 h_T''(\xi_2)}{2} \tag{69}$$

Hence, $t_x - t_*$ has the same order of $\sqrt{n}$.

We are ready to compute the integral in (28). In figure 3, by concavity of $\phi(t)$ on $[t_*, t_x]$, the line segment joining the points $(t_*, \phi(t_*))$ and $(t_x, \phi(t_x))$ is below $\phi(t)$ on the interval $[t_*, t_x]$. Therefore, the integral in (28) is bounded from above by the area of the rectangle between t_1 and t_2, and is bounded below by the area of the trapezoid between t_* and t_x, multiplied by $C_2(n)$. The lower bound is further bounded from below by the area of the triangle between t_* and t_x, multiplied by $C_2(n)$. Hence,

$$\frac{1}{2}C_2(n)\phi(t_*)(t_x - t_*) \leq \int_{t_1}^{t_2} C_2(n)e^{-nI_T(t/n)-\delta t}dt \leq C_2(n)\phi(t_*)(t_2 - t_1) \tag{70}$$

$$\sqrt{n}C_3(n)\phi(t_*) \leq \int_{t_1}^{t_2} C_2(n)e^{-nI_T(t/n)-\delta t}dt \leq nC_4(n)\phi(t_*) \tag{71}$$

where $C_3(n)$ and $C_4(n)$ are $C_2(n)$ modified by a constant multiple. Hence,

$$\int_{t_1}^{t_2} C_2(n)e^{-nI_T(t/n)-\delta t}dt = e^{n\log M_T(-\delta))+o(n)} \tag{72}$$

■

D. First Term

We wish to compute

$$\int_0^{t_1} C_1(n)e^{-nI_T(t/n)}e^{-nI_S(t/n)}e^{-\delta t}dt = \int_0^{t_1} C_1(n)e^{-n(I_T(t/n)+I_S(t/n))-\delta t}dt \tag{73}$$

We start by asking what the maximum of the integrand is and where it occurs. Define

$$I(a) = I_T(a) + I_S(a) \tag{74}$$

By Theorem 6.6 of [3] (page 86), $I(a)$ is the convex conjugate function of

$$\begin{aligned} g(\theta) &= \inf\{\log(M_T(\theta_1)M_S(\theta_2))|\theta_1+\theta_2=\theta\} && (75) \\ &= \inf_{\theta_1}\{\log(M_T(\theta_1)M_S(\theta-\theta_2))\} && (76) \end{aligned}$$

where the domain of g is determined by the domains of $M_T(\theta_1)$ and $M_S(\theta_2)$. If such $g(\theta)$ does not exist, i.e., $g(\theta) = -\infty$ for any θ, it must be true that $I(a) = +\infty$ (check this?), which is not an interesting case because we can just ignore the first term, i.e., (73). Hence, we make the assumption

H6 $g(\theta)$ exists in some neighborhood of 0, and for $a < 1/\mu_1$, $\sup_\theta\{\theta a - g(\theta)\}$ in achieved in the neighborhood.

We then have the conjugate pairs

$$I(a) = \sup_\theta\{\theta a - g(\theta)\} \tag{77}$$

$$g(\theta) = \sup_a\{\theta a - I(a)\} \tag{78}$$

Therefore

$$\begin{aligned} & \sup_t\{-n(I_T(t/n)+I_S(t/n))-\delta t\} \\ = & \sup_t\{-nI(t/n)-\delta t\} \\ = & n\sup_t\{-I(t/n)-\delta t/n\} \\ = & ng(-\delta) && (79) \\ = & n\inf_\theta\{\log(M_T(\theta)M_S(-\delta-\theta))\} && (80) \end{aligned}$$

The integral in (73) is clearly upper bounded

$$\int_0^{t_1} C_1(n)e^{-n(I_T(t/n)+I_S(t/n))-\delta t}dt \le \frac{n}{\mu_1}C_1(n)e^{n\inf_\theta\{\log(M_T(\theta)M_S(-\delta-\theta))\}} \tag{81}$$

Suppose t_o achieves the supremum of $-nI(t/n)-\delta t$. Then, it satisfies

$$-I'(t_o/n)-\delta = -I'_T(t_o/n)-I'_S(t_o/n)-\delta = 0 \tag{82}$$

Then,

$$-h_T^{-1}(t_o/n)-h_S^{-1}(t_o/n)-\delta = 0 \tag{83}$$

where $h_T(\theta)=\frac{M'_T(\theta)}{M_T(\theta)}$ and $h_S(\theta)=\frac{M'_S(\theta)}{M_S(\theta)}$. Since h_T and h_S are both strictly increasing, we can define

$$h(\theta) = (h_T^{-1}+h_S^{-1})^{-1}(\theta) \tag{84}$$

which is also a strictly increasing function. Hence,

$$t_o = nh(-\delta) \tag{85}$$

Lemma 6: t_o exists, is unique, and $0<t_o<t_2=n/\lambda_1$.

Proof: To show $h(-\delta)>0$, it suffices to show that there exists $\epsilon>0$ such that $h(-\delta)\geq\epsilon$. Choose $\epsilon=\min\{h_T(-\delta/2),h_T(-\delta/2)\}$. Since h_T and h_S are both increasing functions, we have $h_T^{-1}(\epsilon)\leq-\delta/2$ and $h_S^{-1}(\epsilon)\leq-\delta/2$. Hence, $(h_T^{-1}+h_S^{-1})(\epsilon)\leq-\delta$. Therefore, $h(-\delta)\geq\epsilon$.

To show the existence of t_o, note that

$$-I'(n\epsilon/n)-\delta = -I'(\epsilon)-\delta = -(h_T^{-1}+h_S^{-1})(\epsilon)-\delta \geq 0 \tag{86}$$

But, because $h_T^{-1}(1/\lambda_1)=0$ and $h_S^{-1}(1/\lambda_1)>h_S^{-1}(1/\mu_1)=0$,

$$-I'(t_2/n)-\delta = -h_T^{-1}(1/\lambda_1)-h_S^{-1}(1/\lambda_1)-\delta < 0 \tag{87}$$

Hence, t_o must exists and is between $[\epsilon,t_2)$, and t_o is unique because I' is strictly increasing. ■

Lemma 7: $t_o\in(0,t_1)$ if and only if $h_T(-\delta)<1/\mu_1$, and hence, $t_o\in[t_1,t_2)$ if and only if $h_T(-\delta)\geq 1/\mu_1$.

Proof: Note that t_o satisfies

$$-I'(t_o/n)-\delta = 0 \tag{88}$$

Since we know $t_o\in(0,t_2)$ and $I'(t/n)$ strictly increases with t, it suffices to show

$$-I'(t_1/n)-\delta < 0 \tag{89}$$

This is true if and only if

$$h_T^{-1}(1/\mu_1)+h_S^{-1}(1/\mu_1) > -\delta \tag{90}$$

But $h_S^{-1}(1/\mu_1) = 0$. Hence, (90) is true if and only if $h_T^{-1}(1/\mu_1) > -\delta$, which is equivalent to $h_T(-\delta) < 1/\mu_1$. ■

Lemma 8:

$$I_S(1/\mu_1) = 0 \tag{91}$$

Proof: By definition,

$$I_S(1/\mu_1) = \sup_\theta \{\theta/\mu_1 - \log M_S(\theta)\} \tag{92}$$

Let θ_2 achieves the supremum. Then,

$$\theta_2 = h_S^{-1}(1/\mu_1) = 0 \tag{93}$$

Hence, $I_S(1/\mu_1) = 0$. ■

1) case 1: $t_o \in (t_1, t_2)$:

Lemma 9: When $t_o \in (0, t_1)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \int_0^{t_1} C_1(n) e^{-n(I_T(t/n)+I_S(t/n))-\delta t} dt = -I_T(1/\mu_1) - \delta/\mu_1 \tag{94}$$

Proof: The proof is similar to that for lemma 3. We first define

$$\phi(t) = e^{-nI(t/n)-\delta t} \tag{95}$$

Then, we have

$$\phi'(t) = e^{-nI(t/n)-\delta t}(-I'(t/n) - \delta) \tag{96}$$

$$\phi''(t) = e^{-nI(t/n)-\delta t}[(-I'(t/n) - \delta)^2 - \frac{1}{n} I''(t/n)] \tag{97}$$

Since $I'(a)$ is an increasing function of a and $-I'(t_o/n) - \delta = 0$, it must be true that $\phi'(t) > 0$ on $(0, t_o)$ and $\phi'(t) < 0$ on (t_o, t_2). Similar to lemma 4, it is easy to show $\phi(t)$ is eventually convex on $(0, t_1)$. Here, we need the assumption that $I''(a)$ is bounded on $a \in [0, 1/\mu_1]$, which is satisfied under the assumption

H7 $I_T''(a)$ and $I_S''(a)$ are continuous on $[0, \infty)$.

which is equivalent to

H7' $h_T'(\theta)$ and $h_S'(\theta)$ are continuous in a neighborhood of 0.

We illustrate the shape of $\phi(t)$ in figure 4. There, we draw the tangent line at the point $(t_1, \phi(t_1))$ whose t-axis intercept is t_x. We have

$$t_1 - t_x = \frac{\phi(t_1)}{\phi'(t_1)} = \frac{1}{I'(1/\mu_1) + \delta} \tag{98}$$

Note that $t_1 - t_x$ is a constant. Hence, for n large enough, t_x falls between 0 and t_1. The integral of $\phi(t)$ on $(0, t_1)$ is bounded from above by the area of the rectangle between $(0, t_1)$ and bounded from below by the area of the triangle between (t_x, t_1). Hence,

$$\frac{1}{2}C_1(n)\phi(t_1)(t_1 - t_x) \leq \int_0^{t_1} C_1(n)e^{-nI(t/n)-\delta t}dt \leq C_1(n)\phi(t_1)t_1 \tag{99}$$

Hence,

$$\int_0^{t_1} C_1(n)e^{-nI_T(t/n)-\delta t}dt = e^{-n(I(1/\mu_1)+\delta/\mu_1)+o(n)} \tag{100}$$

By lemma 8, $I_S(1/\mu_1) = 0$. Hence,

$$I(1/\mu_1) = I_T(1/\mu_1) + I_S(1/\mu_1) = I_T(1/\mu_1) \tag{101}$$

Finally, we have

$$\int_0^{t_1} C_1(n)e^{-nI_T(t/n)-\delta t}dt = e^{-n(I_T(1/\mu_1)+\delta/\mu_1)+o(n)} \tag{102}$$

■

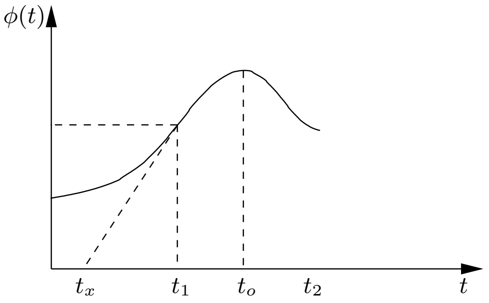

Fig. 4. $\phi(t)$. Case 1: $t_o \in (t_1, t_2)$.

2) case 2: $t_o \in (0, t_1)$:

Lemma 10: When $t_o \in (0, t_1)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \int_0^{t_1} C_1(n)e^{-n(I_T(t/n)+I_S(t/n))-\delta t}dt = g(-\delta) \tag{103}$$

Proof: The proof is very similar to that for lemma 5. We only sketch the proof here. Figure 5 illustrates the situation in this case. First, $\phi''(t) \leq 0$ near t_o, and hence, $\phi(t)$ is concave near t_o. $\phi(t)$ eventually becomes convex as t moves away from t_o. Moreover, for large enough n, all the concave part of $\phi(t)$ falls in $(0, t_1)$. Let $t_x > t_o$ be point where $\phi(t)$ turns from being concave to being convex. It satisfies $\phi''(t) = 0$. We can show that $t_x - t_o$ does not become small as n increases. In fact, under the assumption of **H7** and under the new assumption that

H8 $h''_T(\theta)$ and $h''_S(\theta)$ exist and are continuous in some neighborhood of $-\delta$.

we can show that $t_x - t_o$ is of the order $\sqrt{n}$. We then bound the integral of $\phi(t)$ on $(0, t_1)$ from above by the area of the rectangle between 0 and t_1, and from below by the area of the triangle between t_o and t_x, and get

$$\frac{1}{2}C_1(n)\phi(t_o)(t_x - t_o) \leq \int_0^{t_1} C_1(n)e^{-n(I_T(t/n)+I_S(t/n))-\delta t}dt \leq C_1(n)\phi(t_o)t_1 \tag{104}$$

Because $\phi(t_o) = e^{ng(-\delta)}$, we get the conclusion of the lemma. ■

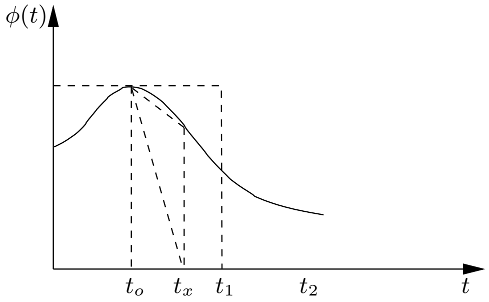

Fig. 5. $\phi(t)$. Case 2: $t_o \in (0, t_1)$.

E. Combining Three Terms

1) case 1: $h_T(-\delta) < 1/\mu_1$: In this case, we need to compare three expressions corresponding to the three terms of (26): $g(-\delta)$, $-(I_T(1/\mu_1) + \delta/\mu_1)$, and $-\delta/\lambda_1$.

Lemma 11: Given $h_T(-\delta) < 1/\mu_1$, we have

$$g(-\delta) \geq -(I_T(1/\mu_1) + \delta/\mu_1) > -\delta/\lambda_1 \tag{105}$$

where

$$g(-\delta) = \inf_\theta\{\log(M_T(\theta)M_S(-\delta - \theta))\} \tag{106}$$

Proof: First let us show $-(I_T(1/\mu_1) + \delta/\mu_1) > -\delta/\lambda_1$. Note that

$$I_T(1/\mu_1) = \theta_o/\mu_1 - \log M_T(\theta_o) \tag{107}$$

where $\theta_o = h_T^{-1}(1/\mu_1)$. Hence, we need to show

$$\log M_T(\theta_o) - \theta_o/\mu_1 - \delta/\mu_1 > -\delta/\lambda_1 \tag{108}$$

which is equivalent to

$$e^{-\theta_o/\mu_1}\mathbf{E}e^{\theta_o T} > e^{-\delta(1/\lambda_1 - 1/\mu_1)} \tag{109}$$

By convexity of the function $x \mapsto e^{\theta_o x}$,

$$e^{-\theta_o/\mu_1}\mathbf{E}e^{\theta_o T} \geq e^{\theta_o(1/\lambda_1 - 1/\mu_1)} \tag{110}$$

We only need to show $\theta_o > -\delta$, which is true because $h_T(-\delta) < 1/\mu_1 = h_T(\theta_o)$ and h_T is increasing.

We next show $g(-\delta) \geq -(I_T(1/\mu_1) + \delta/\mu_1)$. Note that $g(-\delta)$ is the maximum of $-I_T(t/n) - I_S(t/n) - \delta t/n$ over t. Hence

$$\begin{aligned} g(-\delta) &\geq -I_T(t_1/n) - I_S(t_1/n) - \delta t_1/n \\ &= -I_T(1/\mu_1) - I_S(1/\mu_1) - \delta/\mu_1 \\ &= -I_T(1/\mu_1) - \delta/\mu_1 \end{aligned} \tag{111}$$

The last equality is because $I_S(1/\mu_1) = 0$ by lemma 8. ■

2) case 2: $h_T(-\delta) \geq 1/\mu_1$: In this case, we need to compare three expressions corresponding to the three terms of (26): $-(I_T(1/\mu_1) + \delta/\mu_1)$, $\log M_T(-\delta)$, and $-\delta/\lambda_1$.

Lemma 12:

$$\log M_T(-\delta) \geq -\delta/\lambda_1 \tag{112}$$

$$\log M_T(-\delta) \geq -(I_T(1/\mu_1) + \delta/\mu_1) \tag{113}$$

Proof: To show (112), it suffices to show that

$$\mathbf{E}e^{-\delta T} \geq e^{-\delta/\lambda_1} \tag{114}$$

which is true because the function $x \mapsto e^{-\delta x}$ is convex.

To show (113), note that $\log M_T(-\delta)$ is the maximum of $-I_T(t/n) - \delta t/n$ over t. Hence

$$\begin{aligned} \log M_T(-\delta) &\geq -I_T(t_1/n) - \delta t_1/n \\ &= -I_T(1/\mu_1) - \delta/\mu_1 \end{aligned} \tag{115}$$

■

Combining the two cases, we have proven the following theorem.

Theorem 13:

$$\lim_{n\to\infty} \frac{1}{n} \log \int_0^\infty P\{\sum_{i=1}^n T_i \leq t\} P\{\sum_{j=1}^n S_j \leq t\} e^{-\delta t} dt = \begin{cases} g(-\delta) & \text{if } h_T(-\delta) < 1/\mu_1 \\ \log M_T(-\delta) & \text{if } h_T(-\delta) \geq 1/\mu_1 \end{cases} \tag{116}$$

Example 2:

Going back to the exponential case, where T and S and both exponential random variables with mean $1/\lambda_1$ and $1/\mu_1$, respectively. We have $M_T(\theta) = \lambda_1/(\lambda_1 - \theta)$ and $M_S(\theta) = \mu_1/(\mu_1 - \theta)$. We get

$$\log M_T(-\delta) = \log \frac{\lambda_1}{\lambda_1 + \delta} \tag{117}$$

$$g(-\delta) = \inf_\theta \log(\frac{\lambda_1}{\lambda_1 - \theta} \frac{\mu_1}{\mu_1 + \delta + \theta}) \tag{118}$$

By setting $g'(\theta) = 0$, we find that the infimum occurs at $\theta_o = (\lambda_1 - \mu_1 - \delta)/2$, and the infimum is

$$g(-\delta) = \log \frac{4\lambda_1\mu_1}{(\lambda_1 + \mu_1 + \delta)^2} \tag{119}$$

We have already seen in example 1 that the condition $h_T(-\delta) < 1/\mu_1$ becomes $\mu_1 - \lambda_1 < \delta$ in this case.

$$\lim_{n\to\infty} \frac{1}{n} \log \int_0^\infty P\{\sum_{i=1}^n T_i \le t\} P\{\sum_{j=1}^n S_j \le t\} e^{-\delta t} dt = \begin{cases} \log \frac{4\lambda_1\mu_1}{(\lambda_1+\mu_1+\delta)^2} & \text{if } \mu_1 - \lambda_1 < \delta \\ \log \frac{\lambda_1}{\lambda_1+\delta} & \text{if } \mu_1 - \lambda_1 \ge \delta \end{cases} \tag{120}$$

III.

A. *Relaxing the Random Time*

Let Z be a random variable with exponential tail probability. That is, its pdf, $f_Z(t) = e^{-\delta t + o(t)}$ for large t, where $\delta > 0$. We will show that

Theorem 14:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{E} P\{\sum_{i=1}^n T_i \le Z\} P\{\sum_{j=1}^n S_j \le Z\} = \begin{cases} g(-\delta) & \text{if } h_T(-\delta) < 1/\mu_1 \\ \log M_T(-\delta) & \text{if } h_T(-\delta) \ge 1/\mu_1 \end{cases} \tag{121}$$

Proof: For all $\epsilon > 0$, there exists a t_ϵ such that, for all $t > t_\epsilon$,

$$e^{-(\delta+\epsilon)t} \le f_Z(t) \le e^{-(\delta-\epsilon)t} \tag{122}$$

$$\begin{aligned} & \mathbf{E} P\{\sum_{i=1}^n T_i \le Z\} P\{\sum_{j=1}^n S_j \le Z\} \\ = & \int_0^\infty P\{\sum_{i=1}^n T_i \le t\} P\{\sum_{j=1}^n S_j \le t\} f_Z(t) dt \\ = & \int_0^{t_\epsilon} P\{\sum_{i=1}^n T_i \le t\} P\{\sum_{j=1}^n S_j \le t\} f_Z(t) dt + \int_{t_\epsilon}^\infty P\{\sum_{i=1}^n T_i \le t\} P\{\sum_{j=1}^n S_j \le t\} f_Z(t) dt \end{aligned} \tag{123}$$

We will first show that the first term can be ignored when compared with the second term. For large enough n, $t_\epsilon < t_1$. Hence

$$\begin{aligned} & \int_0^{t_\epsilon} P\{\sum_{i=1}^n T_i \le t\} P\{\sum_{j=1}^n S_j \le t\} f_Z(t) dt \\ = & \int_0^{t_\epsilon} C_1(n) e^{-n(I_T(t/n)+I_S(t/n))} f_Z(t) dt \\ \le & \; C_1(n) e^{-n(I_T(t'/n)+I_S(t'/n))} P\{Z \le t_\epsilon\} \\ \le & \; C_1(n) e^{-n(I_T(t'/n)+I_S(t'/n))} \end{aligned} \tag{124}$$

where t' achieves the supremum of $e^{-n(I_T(t/n)+I_S(t/n))}$ on $[0, t_\epsilon]$. Note that

$$\lim_{\theta \to -\infty} M_T(\theta) = \lim_{\theta \to -\infty} \mathbf{E} e^{\theta T} = \mathbf{E}[\lim_{\theta \to -\infty} e^{\theta T}] = 0 \tag{125}$$

The interchange of limit and expectation is justified by dominated convergence theorem. Therefore,

$$I_T(0) = \sup_\theta \{-\log M_T(\theta)\} = +\infty \tag{126}$$

Next, as $n \to \infty$, $t'/n \to 0$. Hence, for any constant $c > 0$, there exists an integer $N(c, \epsilon)$ such that for all $n > N(c, \epsilon)$, $I_T(t'/n) > r$. Hence, for such n,

$$e^{-n(I_T(t'/n)+I_S(t'/n))} \le e^{-nc} \tag{127}$$

That is, $\int_0^{t_\epsilon} P\{\sum_{i=1}^n T_i \le t\} P\{\sum_{j=1}^n S_j \le t\} f_Z(t) dt$ decreases fast than e^{-nc} for any constant c.

We now turn to the second term of (123).

$$\begin{aligned} & \int_{t_\epsilon}^{\infty} P\{\sum_{i=1}^n T_i \le t\} P\{\sum_{j=1}^n S_j \le t\} f_Z(t) dt \\ \le & \int_{t_\epsilon}^{\infty} P\{\sum_{i=1}^n T_i \le t\} P\{\sum_{j=1}^n S_j \le t\} e^{-(\delta-\epsilon)t} dt \\ \le & \int_0^{\infty} P\{\sum_{i=1}^n T_i \le t\} P\{\sum_{j=1}^n S_j \le t\} e^{-(\delta-\epsilon)t} dt \end{aligned} \tag{128}$$

By theorem 13, when $h_T(-\delta+\epsilon) < 1/\mu_1$,

$$\limsup_{n\to\infty} \frac{1}{n} \int_{t_\epsilon}^{\infty} P\{\sum_{i=1}^n T_i \le t\} P\{\sum_{j=1}^n S_j \le t\} f_Z(t) dt \le g(-\delta+\epsilon) \tag{129}$$

and when $h_T(-\delta+\epsilon) \ge 1/\mu_1$,

$$\limsup_{n\to\infty} \frac{1}{n} \int_{t_\epsilon}^{\infty} P\{\sum_{i=1}^n T_i \le t\} P\{\sum_{j=1}^n S_j \le t\} f_Z(t) dt \le \log M_T(-\delta+\epsilon) \tag{130}$$

Because the first term in (123) can be neglected, we have

$$\begin{aligned} &\limsup_{n\to\infty}\frac{1}{n}\int_{t_\epsilon}^{\infty}P\{\sum_{i=1}^{n}T_i\le t\}P\{\sum_{j=1}^{n}S_j\le t\}f_Z(t)dt \\ = &\limsup_{n\to\infty}\frac{1}{n}\int_{0}^{\infty}P\{\sum_{i=1}^{n}T_i\le t\}P\{\sum_{j=1}^{n}S_j\le t\}f_Z(t)dt \end{aligned} \tag{131}$$

For the lower bound,

$$\begin{aligned} &\int_{t_\epsilon}^{\infty}P\{\sum_{i=1}^{n}T_i\le t\}P\{\sum_{j=1}^{n}S_j\le t\}f_Z(t)dt \\ \ge &\int_{t_\epsilon}^{\infty}P\{\sum_{i=1}^{n}T_i\le t\}P\{\sum_{j=1}^{n}S_j\le t\}e^{-(\delta+\epsilon)t}dt \\ = &\int_{0}^{\infty}P\{\sum_{i=1}^{n}T_i\le t\}P\{\sum_{j=1}^{n}S_j\le t\}e^{-(\delta+\epsilon)t}dt \\ &-\int_{0}^{t_\epsilon}P\{\sum_{i=1}^{n}T_i\le t\}P\{\sum_{j=1}^{n}S_j\le t\}e^{-(\delta+\epsilon)t}dt \\ \ge &\int_{0}^{\infty}P\{\sum_{i=1}^{n}T_i\le t\}P\{\sum_{j=1}^{n}S_j\le t\}e^{-(\delta+\epsilon)t}dt \\ &-\int_{0}^{t_\epsilon}P\{\sum_{i=1}^{n}T_i\le t\}P\{\sum_{j=1}^{n}S_j\le t\}dt \end{aligned} \tag{132}$$

For the same reason as before, the second term in (132) decays faster than e^{-ct} for any $c>0$ and can be ignored. Hence, by theorem 13, when $h_T(-\delta-\epsilon)<1/\mu_1$,

$$\liminf_{n\to\infty}\frac{1}{n}\int_{t_\epsilon}^{\infty}P\{\sum_{i=1}^{n}T_i\le t\}P\{\sum_{j=1}^{n}S_j\le t\}f_Z(t)dt\ge g(-\delta-\epsilon) \tag{133}$$

and when $h_T(-\delta-\epsilon)\ge 1/\mu_1$,

$$\liminf_{n\to\infty}\frac{1}{n}\int_{t_\epsilon}^{\infty}P\{\sum_{i=1}^{n}T_i\le t\}P\{\sum_{j=1}^{n}S_j\le t\}f_Z(t)dt\ge \log M_T(-\delta-\epsilon) \tag{134}$$

Because the first term in (123) can be neglected, we have

$$\begin{aligned} &\liminf_{n\to\infty}\frac{1}{n}\int_{t_\epsilon}^{\infty}P\{\sum_{i=1}^{n}T_i\le t\}P\{\sum_{j=1}^{n}S_j\le t\}f_Z(t)dt \\ = &\liminf_{n\to\infty}\frac{1}{n}\int_{0}^{\infty}P\{\sum_{i=1}^{n}T_i\le t\}P\{\sum_{j=1}^{n}S_j\le t\}f_Z(t)dt \end{aligned} \tag{135}$$

Combining the upper and lower bound and since ϵ is arbitrary, we get, when $h_T(-\delta)<1/\mu_1$,

$$\lim_{n\to\infty}\frac{1}{n}\int_{0}^{\infty}P\{\sum_{i=1}^{n}T_i\le t\}P\{\sum_{j=1}^{n}S_j\le t\}f_Z(t)dt=g(-\delta) \tag{136}$$

and when $h_T(-\delta) \geq 1/\mu_1$,

$$\lim_{n\to\infty} \frac{1}{n} \int_0^\infty P\{\sum_{i=1}^n T_i \leq t\} P\{\sum_{j=1}^n S_j \leq t\} f_Z(t) dt = \log M_T(-\delta) \tag{137}$$

Therefore, we have the conclusion of the theorem. ■

IV. LOWER BOUND

Consider a GI/GI/1 queue initially empty at time $t = 0$. Recall that T_n is the interarrival time between customer C_n and customer C_{n+1}, where C_n is the n^{th} customer after time 0, S_n is the service time of C_n, $n = 1, 2, ...$ Let us define A_n be the arrival time, and W_n be waiting time in the queue of customer C_n. Let $N(t)$ be the number of arrivals on the interval $[0, t]$. Suppose a constant β satisfies $0 < \beta < 1/\lambda_1$, and hence, $t := \beta n < n/\lambda_1$. Then, the event $\{A_n \leq t\} = \{\sum_{i=1}^n T_i \leq t\}$ is a large deviation type of event.

Lemma 15: For $t = \beta n$ and for any constant $x > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log P\{A_n \in (t - x, t]\} = -I_T(\beta) \tag{138}$$

Proof: By the theory of large deviation, we know that

$$\lim_{n\to\infty} \frac{1}{n} \log P\{A_n \leq t\} = -I_T(\beta) \tag{139}$$

For large n, we can write

$$P\{A_n \leq t\} = e^{-nI_T(\beta)+o(n)} \tag{140}$$

where $o(n)$ satisfies $o(n)/n \to$ as n tends to infinity. Clearly,

$$P\{A_n \in (t - x, t]\} \leq P\{A_n \leq t\} = e^{-nI_T(\beta)+o(n)} \tag{141}$$

We want to show, for every $\epsilon > 0$, when n is large enough,

$$P\{A_n \in (t - x, t]\} \geq e^{-n(I_T(\beta)+\epsilon)} \tag{142}$$

Suppose this is not the case. Then, there exists an $\epsilon > 0$ such that, for large n,

$$P\{A_n \in (t - x, t]\} \leq e^{-n(I_T(\beta)+\epsilon)} \tag{143}$$

Partition the interval $[0, t]$ into $\lfloor t/x \rfloor = \lfloor \beta n/x \rfloor$ intervals of length x. For any $k = 0, 1, 2, ..., \lfloor \beta n/x \rfloor - 1$,

$$P\{A_n \in (t - x, t]\} \geq P\{A_n \in (t - (k + 1)x, t - kx]\} \tag{144}$$

Hence, we have

$$\begin{aligned}
P\{A_n \le t\} &\le \sum_{k=0}^{\lfloor \beta n/x \rfloor -1} P\{A_n \in (t-(k+1)x, t-kx]\} \\
&\le \sum_{k=0}^{\lfloor \beta n/x \rfloor -1} e^{-n(I_T(\beta)+\epsilon)} \\
&= \lfloor \frac{\beta n}{x} \rfloor e^{-n(I_T(\beta)+\epsilon)} \qquad (145)
\end{aligned}$$

which contradicts (140). ■

Lemma 16: Let X_1, X_2, ..., X_n be identically distributed random variables. Let a be a real number.

$$P\{X_1 + ... + X_j \ge ja \text{ for all } 1 \le j \le n\} \ge \frac{1}{n} P\{X_1 + ... + X_n \ge na\} \qquad (146)$$

Proof: The proof can taken from the proofs for Lemma 3.1 and Theorem 3.3 in [2] (page 46). ■

We will next find a lower bound for the probability that, given n customers arrived on $[0,t]$ to the empty queue, all n customers left by time t. This is the probability that exactly n customers departed the queue on $[0,t]$ leaving the queue empty, given the queue is empty at time 0. This probability can be written as

$$P\{q(t) = 0 \mid N(t) = n, q(0) = 0\} \qquad (147)$$

where $q(t)$ is the number of customers in the queue at t. We will show the following theorem.

Theorem 17: For $t = \beta n$,

$$P\{q(t) = 0 \mid N(t) = n, q(0) = 0\} \ge C(n) P\{\sum_{i=1}^{n} S_i \le t\} \qquad (148)$$

where $C(n)$ is of the order of $e^{o(n)}$.

Proof: A particular way for the n customers to leave the queue by t is that the n^{th} customer did not wait to be serviced and it has departed the queue by t. Therefore,

$$P\{q(t) = 0 \mid N(t) = n, q(0) = 0\} \ge P\{W_n \le 0, A_n + S_n \le t \mid N(t) = n, q(0) = 0\} \qquad (149)$$

Let us define $x = 1/\lambda_1$, and $\rho_1 < \gamma < 1$, where $\rho_1 = \lambda_1/\mu_1$. Since $\mathbf{E}T = 1/\lambda_1$, we get $P\{T_{n+1} \ge x\} > 0$. Since $\gamma x > \rho_1/\lambda_1 = 1/\mu_1$ and $\mathbf{E}S = 1/\mu_1$, we must have $P\{S_n \le \gamma x\} > 0$. We continue to bound the right hand side of (149).

$$\begin{aligned}
& P\{W_n \le 0, A_n + S_n \le t \mid N(t) = n, q(0) = 0\} \\
\ge\ & P\{W_n \le 0, A_n \in (t-x, t-\gamma x], S_n \le \gamma x, T_{n+1} \ge x \mid N(t) = n, q(0) = 0\} \qquad (150)
\end{aligned}$$

As will be seen later, our choice of x makes sure that the probability on the right hand side above does not become zero.

A standard fact from queueing theory is that W_n can be written recursively as

$$W_n = \max(W_{n-1} + S_{n-1} - T_{n-1}, 0) \tag{151}$$

for $n = 2, 3, ...$ Define $U_n = S_n - T_n$, for $n = 1, 2, ...$ We expand the right hand of (151) recursively and use the fact that $W_1 = 0$ since $q(0) = 0$, we get,

$$W_n = \max(U_1 + U_2 + ... + U_{n-1}, U_2 + U_3 ... + U_{n-1}, ..., U_{n-1}, 0) \tag{152}$$

One of the observations is that W_n depends on S_k and T_k for $k = 1, 2, ..., n-1$ and is independent of S_n. Next, $A_n = \sum_{i=1}^{n} T_i$ and is independent of S_n. T_{n+1} is independent of S_n. The following events are equivalent

$$\{N(t) = n\} = \{A_n \leq t, A_{n+1} > t\} = \{A_n \leq t, A_n + T_{n+1} > t\} \tag{153}$$

and hence $\{N(t) = n\}$ is independent of S_n. Finally, $q(0)$ is clearly independent of S_n. Therefore, (W_n, A_n, T_{n+1}) is jointly independent of S_n whether or not conditional on $\{N(t) = n, q(0) = 0\}$. The right hand side of (150) can be written as

$$\begin{aligned}
& P\{W_n \leq 0, A_n \in (t-x, t-\gamma x], S_n \leq \gamma x, T_{n+1} \geq x \mid N(t) = n, q(0) = 0\} \\
= \; & P\{W_n \leq 0, A_n \in (t-x, t-\gamma x], T_{n+1} \geq x \mid N(t) = n, q(0) = 0\} \\
& \cdot P\{S_n \leq \gamma x \mid N(t) = n, q(0) = 0\} \\
= \; & P\{W_n \leq 0, A_n \in (t-x, t-\gamma x], T_{n+1} \geq x \mid N(t) = n, q(0) = 0\} P\{S_n \leq \gamma x\} \qquad (154)
\end{aligned}$$

We have argued that the choice of x makes $P\{S_n \leq \gamma x\}$ a non-zero constant, save $C_1 > 0$. The other probability in (154) can be written as

$$\begin{aligned}
& P\{W_n \leq 0, A_n \in (t-x, t-\gamma x], T_{n+1} \geq x \mid N(t) = n, q(0) = 0\} \\
= \; & P\{W_n \leq 0 \mid N(t) = n, q(0) = 0, A_n \in (t-x, t-\gamma x], T_{n+1} \geq x\} \\
& \cdot P\{A_n \in (t-x, t-\gamma x], T_{n+1} \geq x \mid N(t) = n, q(0) = 0\} \\
= \; & P\{W_n \leq 0 \mid q(0) = 0, A_n \in (t-x, t-\gamma x], T_{n+1} \geq x\} \\
& \cdot P\{A_n \in (t-x, t-\gamma x], T_{n+1} \geq x \mid N(t) = n, q(0) = 0\} \qquad (155)
\end{aligned}$$

In the last step above, we have use the equivalence in (153). We study the two probabilities in (155) separately. By the independence of W_n and T_{n+1},

$$\begin{aligned}
& P\{W_n \leq 0 \mid q(0)=0, A_n \in (t-x, t-\gamma x], T_{n+1} \geq x\} \\
= \; & P\{W_n \leq 0 \mid q(0)=0, A_n \in (t-x, t-\gamma x]\} \\
= \; & P\{\max(U_1+U_2+...+U_{n-1}, U_2+U_3...+U_{n-1}, ..., U_{n-1}, 0) \leq 0 \mid q(0)=0, A_n \in (t-x, t-\gamma x]\} \\
\geq \; & \frac{1}{n-1} P\{U_1+U_2+...+U_{n-1} \leq 0 \mid q(0)=0, A_n \in (t-x, t-\gamma x]\} \qquad (156)
\end{aligned}$$

The last step uses Lemma 16 and the fact the random variables U_1, U_2, ..., U_{n-1} are still identically distributed given $\{q(0)=0, A_n \in (t-x, t-\gamma x]\}$, which can be argued by symmetry.

Now, let us choose an interval $[a, b]$ such that $P\{T_n \in [a, b]\} > 0$ and $b-a < (1-\gamma)x$. This is possible since the CDF of any random variable increases from 0 to 1. Therefore, for any $\epsilon > 0$, one can always find an interval of length ϵ on which the probability is non-zero. Now choose $\eta = a/x + 1$ and $\xi = b/x + \gamma$. Then,

$$\eta - \xi = 1 - \gamma - \frac{b-a}{x} > 0 \qquad (157)$$

Let us suppose A_{n-1} falls into $(t-\eta x, t-\xi x)$ in deriving the following lower bound.

$$\begin{aligned}
& P\{U_1+U_2+...+U_{n-1} \leq 0 \mid q(0)=0, A_n \in (t-x, t-\gamma x]\} \\
= \; & P\{\sum_{i=1}^{n-1} S_i \leq A_{n-1} \mid q(0)=0, A_n \in (t-x, t-\gamma x]\} \qquad (158) \\
= \; & P\{\sum_{i=1}^{n-1} S_i \leq A_{n-1} \mid A_n \in (t-x, t-\gamma x]\} \qquad (159) \\
\geq \; & P\{\sum_{i=1}^{n-1} S_i \leq A_{n-1}, A_{n-1} \in (t-\eta x, t-\xi x) \mid A_n \in (t-x, t-\gamma x]\} \\
\geq \; & P\{\sum_{i=1}^{n-1} S_i \leq t-\eta x, A_{n-1} \in (t-\eta x, t-\xi x) \mid A_n \in (t-x, t-\gamma x]\} \\
= \; & P\{\sum_{i=1}^{n-1} S_i \leq t-\eta x \mid A_{n-1} \in (t-\eta x, t-\xi x), A_n \in (t-x, t-\gamma x]\} \\
& \cdot P\{A_{n-1} \in (t-\eta x, t-\xi x) \mid A_n \in (t-x, t-\gamma x]\} \qquad (160) \\
= \; & P\{\sum_{i=1}^{n-1} S_i \leq t-\eta x\} P\{A_{n-1} \in (t-\eta x, t-\xi x) \mid A_n \in (t-x, t-\gamma x]\} \qquad (161) \\
= \; & P\{\sum_{i=1}^{n-1} S_i \leq t-\eta x\}
\end{aligned}$$

$$\cdot\frac{P\{A_n \in (t-x,t-\gamma x] \mid A_{n-1} \in (t-\eta x,t-\xi x)\}P\{A_{n-1} \in (t-\eta x,t-\xi x)\}}{P\{A_n \in (t-x,t-\gamma x]\}} \tag{162}$$

$$\geq \quad P\{\sum_{i=1}^{n-1} S_i \leq t-\eta x\}$$

$$\cdot\frac{P\{T_n \in [a,b] \mid A_{n-1} \in (t-\eta x,t-\xi x)\}P\{A_{n-1} \in (t-\eta x,t-\xi x)\}}{P\{A_n \in (t-x,t-\gamma x]\}} \tag{163}$$

$$= \quad P\{\sum_{i=1}^{n-1} S_i \leq t-\eta x\}P\{T_n \in [a,b]\}\frac{P\{A_{n-1} \in (t-\eta x,t-\xi x)\}}{P\{A_n \in (t-x,t-\gamma x]\}} \tag{164}$$

From (158) to (159), we use the fact that $S_1,...,S_{n-1}$ and A_{n-1} are independent of $q(0)$ whether or not conditional on A_n. From (160) to (161), we use the fact that S_n is independent of A_{n-1} and A_n. From (162) to (163), we use the fact that $A_n = A_{n-1} + T_n$. When $A_{n-1} \in (t-\eta x, t-\xi x)$ and $T_n \in [a,b]$, $A_n \in (t-\eta x+a, t-\xi x+b)$. By the choice of η and ξ, we have $t-\eta x+a \geq t-x$ and $t-\xi x+b \leq t-\gamma x$. Hence, in this case $A_n \in (t-x, t-\gamma x)$.

Continuing from (164), we first note that, if the distribution of T_n is non-lattice, $P\{A_{n-1} \in (t-\eta x, t-\xi x)\}$ and $P\{A_n \in (t-x,t-\gamma x]\}$ are non-zero for each fixed but large n, and hence, their ratio is non-zero. Second, $\{A_{n-1} \in (t-\eta x, t-\xi x)\}$ and $\{A_n \in (t-x, t-\gamma x]\}$ are either both large-deviation events when $t \leq n/\lambda_1$ or events of constant probability when $t > n/\lambda_1$. In the latter case, $\frac{P\{A_{n-1}\in(t-\eta x,t-\xi x)\}}{P\{A_n\in(t-x,t-\gamma x]\}}$ is a non-zero constant. In the former case, we claim that

$$\frac{P\{A_{n-1} \in (t-\eta x,t-\xi x)\}}{P\{A_n \in (t-x,t-\gamma x]\}} = e^{o(n)} \tag{165}$$

This is immediate by applying Lemma 15. For $t = \beta n$, we have

$$P\{A_n \in (t-x,t-\gamma x]\} = P\{A_{n-1} \in (t-\eta x,t-\xi x)\} = e^{-nI_T(\beta)+o(n)} \tag{166}$$

■

REFERENCES

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

[2] A. Schwartz and A. Weiss, *Large Deviation for Performance Analysis - Queues, Communications and Computing*. Chapman & Hall, 1995.

[3] J. V. Tiel, *Convex Analysis - An Introductory Text*. John Wiley and Sons, 1984.