LARGE DEVIATION THEORY FOR QUEUEING SYSTEMS

Young Wook Park

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Mathematics

APPROVED:

Martin Day

Martin Day, Chairman

Werner Kohler

Werner Kohler

M. Klaus

Martin Klaus

George Terrell

David Russell /M.D.

David Russell

April, 1991
Blacksburg, Virginia

Acknowlegements

I wish to dedicate this thesis to the memory of my father.

I am grateful to God for the many people who have been supportive and helpful throughout the preparation of this thesis and the years that preceded it. Especially, I would like to thank Professor Day for his interest, suggestions, encouragement and for all the lessons imparted. I wish to express my gratitude to my teachers, past and present, and friends who made me finish this thesis. I have been fortunate in having my wife Eun Ju and my mother Hee Suk to whom I owe everything I have.

Table Of Contents

Introduction

A queue, or a waiting line, involves arriving items that wait to be served at the facility, which provides the service they seek. We call this facility "service station" or "station". Arriving items are called "customers". Suppose that the station is a check-out counter at a grocery store. A customer comes in to the queue (waiting line), waiting to be serviced, and after getting served, the customer subsequently leaves the counter. However, the terminology "customer" does not necessarily mean a human customer, it can be a packet of information, waiting in a buffer, to be transmitted to other places in a computer communication system.

A queueing network is a system of service stations, each equipped with a queue, which are interconnected by channels. The numbers of customers in each station, i.e. the queue length, are represented by a multidimensional stochastic process. It is called the "queue length process". The model of our interest is a queueing network which has a finite number N of stations with no bounds on the queue sizes. In this model, no distinction is made between the stochastic characteristics of the customers. If the sequences of stations visited by a customer is denoted by $(A_i)_{i\geq 1}$, the conditional probability law for A_i has the following property:

$$P\{A_{i+1}|A_1, A_2, ..., A_i\} = P\{A_{i+1}|A_i\}, \text{ for any } i \geq 1.$$

Of course, each A_i is the number of a service station, that is $A_i \in 1, 2, ..., N$, or $A_i = 0$ to represent being in the exterior of the network. In each station, there is a server who provides customers in this station with services and the sever can serve only one customer at a time. Services are provided in the rule of first-come-first-serve. The times for completing serving each customer in the station are independent of each other and distributed according to the exponential law having a parameter which describes the efficiency of service of the station. Arrival times between customers from the exterior to a station are also independent of each other and follow the exponential distribution. Service patterns of stations and arrival patterns to stations from the exterior are independent of each other. The queue length process of this system is a N dimensional Markov jump process.

This model, known as a Jackson network, is well-known due to the interesting property of having a stationary distribution in 'product' form. It is widely used to model complex computer systems and data transmission systems. Even though a lot of work has been done on this model, many interesting and important features, especially of transient behavior, have not been investigated. For example, in a system with finite queues, the probability of the system starting from an empty state and going to a state in which at least one queue reaches a certain large number and the first passage time of this event are two of those.

Recently, it has been reported in [3], [11], [15], and [23] that large deviation theory has been used as a tool to analyze the transient behavior of stochastic systems, especially communication networks and queueing networks. Even though it has been used widely in engineering research, the large deviation theory for queueing systems has not been fully established, yet. Therefore, applications of the large deviation theory depend on many conjectures. This thesis concerns the development of large deviation principles (LDP) for Jackson networks.

Large deviation principles for a certain class of Markov processes have been well established by Wentzel [5] under carefully designed conditions by Wentzel [5]. However, many interesting Markov processes do not belong to the class which Wentzel assumed. Their probability structures violate some of the conditions which are assumed by Wentzel. The queue length process of a Jackson network is one of them. The goal of this research is to establish large deviation principles for appropriate classes of queueing processes.

Chapter 1 will briefly introduce the large deviation theory and Wentzel's result for a special class of Markov processes. In chapter 2, we will carefully specify the class of processes to be studied. The $M/M/1$ model (the simplest model of Jackson network which has only one station) will be used to illustrate the main reason that we can not apply Wentzel's result directly to Jackson networks. We will discuss the current research which are related to the large deviation theory for queueing networks and describe the Skorohod problem approach which will be developed in this study. We will sketch this approach in particular for $M/M/1$ queueing system.

Chapter 3 will define a more general class of Jackson networks for which a Skorohod problem approach can be used. The Skorohod problem with oblique reflexion boundary will be explained in a general setting. We will develop the Skorohod problem representation for our class of systems. This allows us to obtain the LDP for queueing processes of our class by applying the "contraction principle" of general large deviation theory together with Wentzel's result for Markov processes.

In Chapter 4, we will show that the LDP obtained in our approach is the same as that of Dupuis [6], which was developed with a completely different approach. The Skorohod problem approach that we present in this thesis does not work in a more general class of Jackson networks. The reason for this will be explained.

Chapter 1. Large Deviation Principles of Markov Processes

In this chapter we shall consider theorems on the asymptotics of probabilities of large deviations for Markov (random) processes.

1.1 Introduction to Large Deviations.

In this section, we introduce the general large deviation theory on a set of measures briefly;
Let X be a metric space.

DEFINITION 1.1. *A parameterized family of probability measures, $\{P^\epsilon\}$, on X is said to obey the large deviation principle with the rate function $I(.)$ if there is a function $I(.)$ from X to $[0,\infty]$ satisfying:*

(1) $I(.)$ is lower semicontinuous.

(2) For each $c < \infty$, the set $\{x$: $I(.) \leq c\}$ is a compact set in X.

(3) For each closed set $F \subset X$,
$\limsup_{\epsilon\to 0} \epsilon log P^\epsilon(F) \leq -\inf_{x\in F} I(x)$.

(4) For each open set $G \subset X$,
$\liminf_{\epsilon\to 0} \epsilon log P^\epsilon(G) \geq -\inf_{x\in G} I(x)$.

It follows that if A is a Borel subset of X such that $inf_{x\in A^o} I(x) = inf_{x\in A} I(x) = inf_{x\in A^c} I(x)$, then $lim_{\epsilon\to 0}\epsilon log P^\epsilon(A) = -inf_{x\in A} I(x)$. In other words,

$$P^\epsilon(A) \approx \exp[-(1/\epsilon)\inf_{x\in A} I(x)],$$

up to logarithmic equivalence. It tells that unless A contains an element x_o such that $I(x_o) = 0$, $P^\epsilon(A)$ converges to 0 exponentially fast at the rate which $I(\cdot)$ and A determine.

In the following, we consider a set of probability measures, $\{P^\epsilon\}$ as an example, which will converge weakly to a probability measure which degenerate at some point

x_0 in X, as $\epsilon \to 0$.

EXAMPLE 1.2.

$X_1, X_2, ..., X_n$ are n independent random variables which have a common normal distribution with $E(X_i) = 0$, $Var(X_i) = 1$. Let $P_n = P \circ (\frac{S_n}{n})^{-1}$, $S_n = \sum_{i=1}^{n} X_i$. By the law of large numbers, $\frac{S_n}{n} \to 0$ in probability. Therefore, P_n converges to δ_0 weakly, where δ_0 is a dirac delta function at x=0. And $P_n(A) = \sqrt{\frac{n}{2\pi}} \int_A exp[-\frac{nx^2}{2}]dx$ converges to 0 exponentially fast if 0 is not in A and the decreasing rate is dependent on A. In fact, one can easily check that $\{P_n\}$ obey the large deviation principle with rate function:

$$I(x) = \frac{x^2}{2}.$$

That is, $lim_{n\to\infty} \frac{1}{n} log P_n(A) = -inf_{x\in A} \frac{x^2}{2}$ for any interval, A. This implies that $P_n(A) \approx exp(-n \inf_{x\in A} \frac{x^2}{2})$.

The first general limit theorems for probabilities of sums of independent random variables as Example 1.2 are contained in Cramer's paper [4]. The rate function for probabilities of sums of independent random variables can be formulated in terms of the Legendre transformations of some convex functions connected with the exponential moments of the random variables under the assumption that exponential moments are finite.

Before starting the next section, we define notations and metrics on the spaces which we consider later on.

NOTATION.

Let $\alpha \in R^r$.

$|||\alpha||| = sup_{i=1,..,r} |\alpha_i|$.

$||\alpha|| = \sqrt{\sum_{i=1,..,r} \alpha_i^2}$.

$\alpha \geq 0$ means that $\alpha_i \geq 0$, $\forall i = 1, \cdots, r$.

$\alpha > 0$ means that $\alpha_i > 0$, $\forall i = 1, \cdots, r$.

The inner product of x and y is denoted by $\langle x, y \rangle$.

Let G be a set.

G^o denotes the interior of G.

G^c denotes the closure of G.

For a positive scalar c, $cG = \{cx : x \in G\}$.

Let A be a matrix.

A_i denotes the i^{th} column of the matrix A.

A^i denotes the i^{th} row of the matrix A.

A^T is the transpose of A.

Let ρ_T be the uniform metric on $F^r[0,T] = \{f : [0,T] \to R^r\}$, i.e. if f and g are in $F^r[0,T] = \{f : [0,T] \to R^r\}$,

$$\rho_T(f,g) = \sup_{t \in [0,T]} |||f(t) - g(t)|||.$$

Let T be a positive real number.

$D^r[0,T]$ is a space of right continuous function $x(.)$: $[0,T] \to R^r$, which has left limits at all $t \in [0,T]$.

$D^{r(+)}[0,T]$ is a space of nondecreasing, right continuous function $x(.)$: $[0,T] \to R^r$, which has left limits at all $t \in [0,T]$.

$D^r_{(+)}[0,T]$ is a space of nonnegative, right continuous function $x(.)$: $[0,T] \to R^r$, which has left limits at all $t \in [0,T]$.

1.2 Legendre Transformation.

The Legendre transformation is important in formulating LDP of Markov processes because the family of Markov processes we are going to consider can be viewed as generalizations of the scheme of summing independent random variables; the constructions used in the study of large deviations for Markov processes generalize constructions encountered in the study of sums of independent terms. Therefore, Legendre transformation turns out to be essential in our study. We consider this transformation and its application to families of measures in finite-dimensional spaces.

Let $H(\alpha)$ be a function of an $r-$dimensional real vector, assuming its values in $(-\infty, \infty]$ and not identically equal to $+\infty$. Suppose that H(α) is a convex and lower semicontinuous function on R^r. $L(\beta)$ is said to be the Legendre transformation of H(.) at $\beta \in R^r$ if

$$L(\beta) = sup_{\alpha \in R^r}[\langle \alpha, \beta \rangle - H(\alpha)],\ \beta \in R^r. \tag{1.1}$$

We will review a number of properties of the Legendre transformation below. A thorough treatment of these and other properties of the Legendre transformation can be found in Rockafellar [24].

First, $L(.)$ is also a convex and lower semicontinuous function, assuming values in $(-\infty, +\infty]$ and not identically equal to ∞. The inverse of the Legendre transformation is itself:

$$H(\alpha) = sup_{\beta \in R^r}[\langle \alpha, \beta \rangle - L(\beta)],\ \alpha \in R^r. \tag{1.2}$$

The functions L and H coupled by relations (1.1) and (1.2) are said to be conjugate. If H(0) =0, then by (1.1), $L(\beta) \geq \langle 0, \beta \rangle - H(0) = 0$ for all β. Therefore,

$$\text{if } \ H(0) = 0, \text{ then } L(\beta) \geq 0, \text{ for all } \beta. \tag{1.3}$$

It turns out that $L(\beta) \to \infty$ as $||\beta|| \to \infty$ if and only if $H(\alpha) < \infty$ in some neighborhood of $\alpha = 0$. For functions H, L which are analytic, we can find the solutions $\alpha = \alpha(\beta)$ to the equations $\nabla H(\alpha) = \beta$ for each β and $L(\beta)$ is determined from the formula

$$L(\beta) = \langle \alpha(\beta), \beta \rangle - H(\alpha(\beta)). \tag{1.4}$$

Moreover, we have $\alpha(\beta) = \nabla L(\beta)$. If one of the functions conjugate to each other is more than twice continuously differentiable, and the Jacobian matrices are positive definite everywhere, then the other function has same the smoothness and the Jacobian matrices at corresponding points in the relation of (1.4) are inverses of each other:

$$(\frac{\partial^2 L(\beta)}{\partial\beta_i\partial\beta_j}) = (\frac{\partial^2 H}{\partial\alpha_i\partial\alpha_j}(\alpha(\beta)))^{-1}. \tag{1.5}$$

1.3 Markov Processes and Their Generators.

As we mentioned in the introduction of this chapter, the rate function for probabilities of sums of independent random variables can be formulated in terms of the Legendre transformations of some convex functions connected with the exponential moments of the random variables. The role which the exponential moments play in the case of sums of variables belongs to (infintesimal) generators of Markov processes. The following is the definition of the generator.

Let Y(t) be a time homogeneous Markov process in R^r with a set of transition function,$\{P_{x,t}\}$,

$$P_{x,t}(\cdot) = Pr[Y(t) \in \cdot \mid Y(0) = x].$$

In this study, we specifically look at Markov jump processes whose infinitesimal generator has the following form.

$$Af(x) = \int_{R^r-\{0\}} [f(x+\beta) - f(x)]\mu_x(d\beta), \tag{1.6}$$

where μ_x is a measure on $R^r - \{0\}$ such that

$$\int_{R^r-\{0\}} |\beta|^2\mu_x(d\beta) < \infty.$$

It follows from this that $Y^\epsilon(t) = \epsilon Y(t/\epsilon)$ (throughout this research, we use this rescaling method.) has the following generator.

$$A^\epsilon f(x) = \epsilon^{-1}\int_{R^r-\{0\}} [f(x+\epsilon\beta) - f(x)\mu_x](d\beta). \tag{1.7}$$

1.4 Rate Functions of Markov Processes.

Let

$$H(x,\alpha) = \int_{R^r-(0)} [exp(\sum_{i=1}^{r} \alpha_i\beta_i) - 1]\mu_x(d\beta). \tag{1.8}$$

Wentzel [30] points out that, in general, for a parameterized family of Markov processes with generators A^ϵ, the appropriate $H(\cdot,\cdot)$ is described by $A^\epsilon exp(\sum_{i=1}^{r} \alpha_i x_i)$ $= \epsilon^{-1}H(x,\epsilon\alpha)\ exp(\sum_{i=1}^{r} \alpha_i x_i)$ (see Chapter 5 in Freidlin and Wentzel [12]).

Notice that H is a convex and analytic function with respect to α and $H(x,0) = 0$.

For each x, let $H(x,\alpha)$ be the function in (1.8) and $L(x,\beta)$ is the Legendre transformation of $H(x,\alpha)$ in the second variable. Now, let's define the most important functional in this study;

$$I_{0,T}(x,\phi) = \int_0^T L(\phi(t),\phi'(t))dt, \tag{1.9}$$

if ϕ is a r-dimensional absolutely continuous function , its integral is convergent, and $\phi(0) = x$. Otherwise, it is equal to ∞.

In order that $I_{[0,T]}(x,\phi)$ be the rate function for the family of the rescaled processes $\{Y^\epsilon(t), P_x^\epsilon\}$, it is necessary to impose some restriction on this family. We formulate them in terms of the functions H and L.

(1.10) There is a function $\overline{H(\alpha)}$ which is everywhere a finite nonnegative convex function with $\overline{H(0)} = 0$ and $H(x,\alpha) \le \overline{H(\alpha)}$ for all x, α.

(1.11) For any $R > 0$, there exist $M, m > 0$ such that $L(x,\beta) \le M$, $||\nabla L(x,\beta)|| \le M$, $\sum_{i,j}(\frac{\partial^2 L}{\partial\beta_i\partial\beta_j})(x,\beta)c_ic_j \ge m\sum_i {c_i}^2$ for all $x, c \in R^r$ and all β, $||\beta|| < R$.

(1.12)

$$sup_{|x-x'|<\delta}sup_\beta \frac{L(x',\beta) - L(x,\beta)}{1 + L(x,\beta)} \to 0 \quad as\ \delta \to 0.$$

We can summarize the previous discussion with the following theorem and it provides the foundation of the general scheme of this study;

THEOREM 1.3 [WENTZEL]. *Suppose that the function $H(x,\alpha)$ and $L(x,\beta)$ satisfy the conditions (1.10) - (1.12) and the function $I_{[0,T]}(x,\phi)$ defined by the formula (1.9).*

Then $I_{[0,T]}(x,\phi)$ is the rate function for the family of rescaled processes $(Y^\epsilon(t), P_x^\epsilon)$ on $D^r[0,T]$ when $D^r[0,T]$ is equipped with the topology generated by the uniform metric ρ_T. That is,

(1.13) For each $c < \infty$, the set $\cup_{x\in A}\{\phi : I_{[0,T]}(x,\phi) \le c\}$ is a compact set in $D^r[0,T]$ for any compact set $A \subset R^r$, for all $0 < T < \infty$.

(1.14) For each closed set $F \subset D^r[0,T]$,

$$\limsup_{\epsilon\to 0} \epsilon log P[Y^\epsilon(.) \in F | Y^\epsilon(0) = x] \le -\inf_{\phi\in F} I_{[0,T]}(x,\phi).$$

(1.15) For each open set $G \subset D^r[0,T]$,

$$\liminf_{\epsilon\to 0} \epsilon log P[Y^\epsilon(.) \in G | Y^\epsilon(0) = x] \ge -\inf_{\phi\in G} I_{[0,T]}(x,\phi).$$

The proof of this theorem is given in Wentzel [30].

REMARK: This result holds for a more general class of Markov processes which allows diffusion as well as random jumps. However, since we focus on Markov jump processes, we simply this result for our own purpose.

If there exists the rate function, $I_{[0,T]}(x,\phi)$, the family of rescaled processes $\{Y^\epsilon(t), P_{x,t}^\epsilon\}$ are said to obey a large deviation principle with rate function, $I_{[0,T]}$.(For convenience, we may say "LDP of the process Y".) It implies that if $B \subset D^r[0,T]$ and $inf_{\phi\in B^o} I_{[0,T]}(x,\phi) = inf_{\phi\in B^c} I_{[0,T]}(x,\phi)$, then $P[Y^\epsilon(.) \in B | Y^\epsilon(0) = x] \approx exp[-\frac{1}{\epsilon}\inf_{\phi\in B} I_{[0,T]}(x,\phi)]$, $(UTLE)$.

Moreover, if there exists a unique ϕ^* in B such that $I_{[0,T]}(x,\phi^*)$=$\inf_{\phi\in B} I_{[0,T]}(x,\phi)$,

$$P[Y^\epsilon(.) \in B | Y^\epsilon(0) = x] \approx P[Y^\epsilon(.) \in \phi_\delta^* | Y^\epsilon(0) = x],$$

where ϕ_δ^* is an arbitrary small tube, containing ϕ^*, i.e. ϕ_δ^*= $\{\phi : \rho(\phi,\phi^*) < \delta\}$ for an arbitrary small positive δ. In this sense, we call ϕ^* "the most likely trajectory" of the event B.

Many interesting processes do not fit into the class of Markov processes for which this general theory has been developed. The queue length process in a queueing system, especially a Jackson networks as considered in this study, is one of them. The main goal of this research is to find the LDP's of queue length processes of Jackson networks.

Our basic approach will be to find a Markov process which satisfies the conditions (1.10)-(1.12) under $Y^\epsilon(t) = \epsilon Y(t/\epsilon)$ and a continuous mapping through which we can transform the process into the process of our interest. Then we will push the LDP of the original process through this map to get the LDP of the process of our interest. The following theorem is the main vehicle of "pushing through",usually called the "contraction principle".

THEOREM 1.4. *X is a metric space. $\{P^\epsilon\}$ satisfies the large deviation principle with a rate function $I(.)$. θ is a continuous map from $X \to Y$ where Y is another metric space. If we define $Q^\epsilon = P^\epsilon\theta^{-1}$, then $\{Q^\epsilon\}$ satisfies the large deviation principle with a rate function $J(.)$, defined by $J(y) = inf_{x:\theta(x)=y} I(x)$ (the minimum over the empty set is set to be equal to ∞).*

REMARK: The proof can be found in Theorem 3.1 in [12].

1.5 The Birth and Death Process.

The birth and death process is an important example to which Theorem 1.3 does apply. In this section we will determine its rate function as given by (1.9).

The birth-death process $X(t)$ is the time homogeneous Markov jump process on the integer state space with jump rates $\lambda_1,\ \lambda_2 > 0$:

$$P[X(h) = j|X(0) = i] = \begin{cases} \lambda_1 h + o(h), & j = i+1 \\ \lambda_2 h + o(h), & j = i-1 \\ 1 - \lambda_1 h - \lambda_2 h + o(h), & j = i \\ 0, & o.w.. \end{cases}$$

The rescaled process $X^\epsilon(t) = \epsilon X(\frac{t}{\epsilon})$ has small time increment probability related to those of X as follows:

$$P_x^\epsilon(B) = P[X^\epsilon(t+h) \in B | X^\epsilon(t) = x]$$

$$
\begin{aligned}
&= P[\epsilon X(\frac{t+h}{\epsilon}) \in B | \epsilon X(\frac{t}{\epsilon}) = x] \\
&= P[X(\frac{t+h}{\epsilon}) \in \frac{B}{\epsilon} | X(\frac{t}{\epsilon}) = \frac{x}{\epsilon}] \\
&= P[X(\frac{h}{\epsilon}) \in \frac{B}{\epsilon} | X(0) = \frac{x}{\epsilon}] \qquad (1.16)
\end{aligned}
$$

It follows that the infinitesimal generator,A^ϵ, of $X^\epsilon(t)$ is given by

$$
\begin{aligned}
A^\epsilon f(x) =& lim_{h\to 0} \frac{E[f(X^\epsilon(h)) - f(X^\epsilon(0)|X^\epsilon(0) = x]}{h} \\
=& [f(x+\epsilon) - f(x)]\lambda_1(\frac{1}{\epsilon}) + [f(x-\epsilon) - f(x)]\lambda_2(\frac{1}{\epsilon}).
\end{aligned}
$$

Thus $A^\epsilon f(x)$ has the form ,defined in (1.7), i.e.,

$$
A^\epsilon f(x) = \epsilon^{-1} \int_R [f(x+\epsilon\beta) - f(x)]\mu(d\beta),
$$

where $\mu(1) = \lambda_1$ and $\mu(-1) = \lambda_2$. In order to check that (1.3) applies to this simple Markov jump process, we need to verify (1.10)-(1.12) for the associate L and H. The definition (1.8) of H(.) in this example,

$$
\begin{aligned}
H(x,\alpha) =& \int_R (exp(\alpha\beta) - 1)\mu(d\beta) \\
=& \lambda_1(exp(\alpha) - 1) + \lambda_2(exp(-\alpha) - 1). \qquad (1.17)
\end{aligned}
$$

On may readily calculate that

$$
L(x,\beta) = \beta log \frac{\beta + \sqrt{\beta^2 + 4\lambda_1\lambda_2}}{2\lambda_1} - \sqrt{\beta^2 + 4\lambda_1\lambda_2} + \lambda_1 + \lambda_2; \qquad (1.18)
$$

see Freidlin and Wentzel [12]. Notice that H does not depend on x and neither does L and $L(x, \lambda_1 - \lambda_2) = 0$. Conditions (1.10), and (1.12) are trivially satisfied. Since L is analytic and independent of x, there exist M, $m > 0$ such that $L(x,\beta) \le M$, $\|\nabla L(x,\beta)\| \le M$ for all β, $|\beta| < R$. Since

$$
(\frac{\partial^2 H}{\partial \alpha^2})(x,\alpha) = \lambda_1 exp(\alpha) + \lambda_2 exp(-\alpha) \ > 0, \ \forall\alpha.
$$

By (1.5), (1.11) is satisfied. Therefore, $I_{[0,T]}(\phi)$, defined in (1.9), is the rate function of the large deviation principle for the birth and death process. In particular when $x = 0$, for appropriate sets of trajectories $G \subset D[0,T]$, the LDP for $X^\epsilon(.)$ says that

$$
P[X^\epsilon(.) \in G | X^\epsilon(0) = x] \approx exp[-\frac{1}{\epsilon} \inf_{\phi\in G} I_{[0,T]}(x,\phi)], \ as \ \epsilon \to \ 0,
$$

where $I_{[0,T]}(x, \phi)$ measures the deviation of ϕ from the limiting process $x(t) = (\lambda_1 - \lambda_2)t$ which is the solution to the equation $I_{[0,T]}(0, \phi) = 0$, in a sense which is appropriate for calculating estimates of $P[X^\epsilon \approx \phi(.)]$.

Chapter 2. Queueing Processes

2.1 Formulation.

As we briefly discussed in the introduction, the queueing network is a system of N service stations equipped with queues which are interconnected through channels. We are interested in a special type of Jackson network: in which there is exactly one channel out of any station and one channel connects one station or the (network) exterior to one station or the exterior. Obviously, there is no channel which connects the exterior to the exterior. We can label the stations 1 through N so that if there is a sequence of channels from i station to j station, then $i < j$. This assumption implies that there is no feedback in the system. This describes the topological structure of this system.

Regarding the probability structure of this system, we will assume that the service times at each station and interarrival times between customers from the exterior to stations are independent and exponentially distributed with parameters that are fixed and do not depend on the status of the system. We will refer to such queueing systems as "One Channel from One Station" (OCOS) systems. Let $X^i(t)$ denote the number of customers in the i^{th} station at time t and $X(t)=(X^1(t), \cdots, X^N(t))^T$. We call it the queue length process of this system.

Define

$\mathbb{Z} = \{$ integers$\}$,

$\mathbb{Z}_N = \{1, 2, \cdots, N\}$,

$\overline{K} = \{subsets\ of\ K\}$, where $K \subset \mathbb{Z}_N$, and

$B(x) = \{i : x_i = 0\}$, $x \in R^r$.

Each channel in the system will be associated with an element v of $\{0, 1, -1\}^N$. For a channel which connects i station to j station, then $v_i = 1$, $v_j = -1$ and $v_k = 0$, $\forall k \neq i, j$. For a channel from i station(the exterior) to the exterior(i station), then v_i is equal to -1(1) and the rest are zeros. Let V be the collection of v which represents a channel in the given system. Due to the topological structure of this system, for each $i = 1, ..., N$, there is only one $v \in V$ with $v_i = -1$. We write it $v(i)$. V_K is a set of $v(i)$ for i in K.

Since there is a service station or the exterior at the beginning of a channel and the distribution of service times at the station or interarrival times follows an exponential distribution with a parameter, we can assign a positive parameter λ_v to a corresponding $v \in V$.

Define $\lambda_v(x)$, for each v,

$$\lambda_v(x) = \begin{cases} 0, & i \in B(x) \ and \ v_i = -1 \\ \lambda_v, & o.w.. \end{cases}$$

These are the state dependent jump rates for the associated queue length process, $X(t)$.

$X(t)$ is a time homogeneous Markov process, due to the memoryless property of the exponential distribution. That is, for any event A,

$$\begin{aligned} P[X(t+h) \in A | X(u),\ 0 \le u \le t] &= P[X(t+h) \in A | X(t)] \\ &= P[X(h) \in A | X(0)]. \end{aligned}$$

Moreover, for sufficiently small $h > 0$, the small time increment conditional probability for each initial condition can be expressed in the following manner,

$$P[X(h) - X(0) = v | X(0) = x] = \begin{cases} \lambda_v(x)h + o(h), & v \in V,\ v \neq 0 \\ 1 - \sum_{v \in V} \lambda_v(x)h + o(h), & v = 0 \\ 0, & o.w.. \end{cases}$$

Note that it depends on the initial state x.

Consider the asymptotic analysis of the rescaled system, $\epsilon X(t/\epsilon)$. If every element of x is strictly positive, then the small time increment conditional probability of $\epsilon X(t/\epsilon)$ is roughly independent of x in the sense that as long as every element of x is positive, the probabilities are the same. However, when at least one of its elements becomes 0, then there is an abrupt change in its probability structure, due to the nonnegativity of queue length processes. This discontinuity *w.r.t.* x is the new feature involved in developing large deviation theory for queueing processes.

In section 2.2, we survey the existing researches which are related to the large deviations of queueing systems and describe their successes and limitations. In section 2.3, we introduce, a bit heuristically, the main idea of our approach with a simple example, the $M/M/1$ queueing system.

2.2. Related Research.

We mentioned in the introduction that the motivation for developing a large deviations theory for the processes that model queues arises from the desire to find useful estimates of the probabilities of rare transient events in queueing systems. Several approaches have been made to obtain information about such events. For instance, we might be concerned with the probability that the queue length process X of a queueing system of N stations avoids some specified set G on $[t, T]$;

$$Pr[X(s) \notin G \ for \ some \ t \leq s \leq T | X(t) = x]. \tag{2.1}$$

One such $G \subset R^{N(+)}$ of interest might be $\{x \in R^{N(+)} : Max_{i=1,\cdots,n} x_i \leq N\}$. In lieu of an exact representation one may seek various asymptotic expressions instead. Suppose a scaling parameter $\epsilon > 0$ is introduced into (2.1):

$$Pr[X(s) \notin (1/\epsilon)G \ for \ some \ (t/\epsilon) \leq s \leq (T/\epsilon) | X(t/\epsilon) = x/\epsilon], \tag{2.2}$$

where $\alpha G \equiv \{\alpha x : x \in G\}$. Let $v^\epsilon(x)$ denote the quantity defined by (2.2). So far, we can try to obtain effective information about (2.1) by studying asymptotic estimates of (2.2) for sufficiently small ϵ.

There are three quite different approaches in this area. One method depends on a WKB type expression in ϵ:

$$v^\epsilon(x) = exp(-u(x)/\epsilon)(v_o + \epsilon v_1 + \cdots + \epsilon^m v_m + o(\epsilon^m)).$$

Expansions of this type have been obtained for several cases of queueing systems, for instance see [18], [19], [20], and [21]. This approach depends on perturbation methods for partial differential equations which are induced from the dynamics of the processes. In the papers above, specific queueing models with simple structures, (M/M/1, M/G/1, or M/G/2) were chosen and specific features of the system are studied extensively. However, to our knowledge, none of them deal with multidimensional queueing networks. The non-nonnegativity constraints on queue sizes become boundary value constraints for the partial differential equations of the queueing systems. These boundary value problems cause difficulties in extending this approach for multidimensional queueing networks like Jackson networks. At the present time, no systematic way to find such quantities as (2.2) has been established via this approach.

Instead of $v^\epsilon(x)$, the other two approaches focus on

$$u(x) = lim_{\epsilon\to 0} - \epsilon \; ln \; v^\epsilon(x). \tag{2.3}$$

This $u(x)$ is closely related to the rate function of this process. That is, if there exists a rate function $I_{[0,T]}(x,\cdot)$ for the queue length process X with $X^\epsilon(t) = \epsilon X(t/\epsilon)$, then the $u(x)$ can be determined from it. Threrefore, the center of interest is placed on the LDP of the process X and its rate function. On the other hand, if $u(x)$ can be found by some method for an appropriate family of G, then under some additional hypotheses the rate function $I_{[0,T]}(x,\cdot)$ for the queueing system can be established.

One method for studying (2.3) uses the notion of "viscosity solution", due to M. Crandall and P. -L. Lions [5]. The application of results for the theory of viscosity solutions to finding the LDP of queue length processes begins with the work of Dupuis et. al. [6]. In this paper, they consider a specific Jackson network with two stations which allows routing and looping. Based on the viscosity solution approach, a formula for the quantity $u(x)$ was discovered. They then determined what the LDP of the system would have to be from the formula for $u(x)$. Moreover, under certain assumptions, Dupuis [9] provided the general scheme for finding the rate function of the process that models a general queueing system via this approach. For a given particular Jackson network, the assumed conditions might be checked. However, at the present time, there is no general technique for proving the assumptions in the case of a general Jackson network.

In contrast to the results available by viscosity solution methods, the other approach which is more probabilistic has so far given only an upper bound, but in a much more general setting [7]. The main idea of this approach is based on the fact that

$$exp[1/\epsilon\{< X^{\epsilon}(t) - X^{\epsilon}(s), \alpha > - \int_{s}^{t} H(X^{\epsilon}(u), \alpha) du\}],$$

is a martingale in t for $0 \leq s \leq t$ with an arbitrary $N-$ dimensional vector α. The main difficulty in finding an lower bound arises from the discontinuity problem on the boundary which we described in the section 2.1.

There is presently no established method to find the LDP of the process which models a general queueing system. In our research, we find a systematic way to establish the LDP of Jackson networks with a special structure, which we will call OCOS systems. Basically, our approach depends on the work of Wentzel et. al. [12]. The new feature of our research is our approach to solving the discontinuity problem on the boundary, which we explained in the above. We call our approach the"Skorohod problem approach", because we use the mapping which originated from the solution of the Skorohod problem which we will describe in Theorem 2.1 and Chapter 3 and Chapter 4. This approach owes a lot to the groundwork that Dupuis et. al. [8] lay. However, our approach is not successful in Jackson networks which have a more general structure than the OCOS structure which we assume below. We will discuss the method's shortcomings in Section 3.4.

2.3 Skorohod Problem Approach in the $M/M/1$ Queue.

The goal of this study is to find a rate function for the large deviation principle for the OCOS Jackson networks we described in section 2.1. In this section, we will illustrate the main idea of our research using the M/M/1 queue, which is described in the following.

Time intervals between arrivals are independent and exponentially distributed with a positive parameter λ_1 . Service times are also independent and exponentially distributed with a positive parameter λ_2. The interarrival times and service times are also independent. The queue length process of this system is a 1-dimensional time homogeneous Markov jump process with the nonnegative integers as the state space. Therefore,

$$\begin{aligned} & P[X(t+h) - X(t) = j | X(t) = i,\ X(s),\ 0 \le s \le t] \\ & = P[X(t+h) - X(t) = j | X(t) = i] \\ & = P[X(h) - X(0) = j | X(0) = i]. \end{aligned}$$

In particular, when i is strictly positive,

$$P[X(h) = j | X(0) = i] = \begin{cases} \lambda_1 h + o(h), & j = i+1 \\ \lambda_2 h + o(h), & j = i-1 \\ 1 - \lambda_1 h - \lambda_2 h + o(h), & j = i \\ 0, & o.w.. \end{cases} \tag{2.4}$$

However, when the station is empty, the conditional probability distribution of small time increment is not same as the case above.

$$P[X(h) = j | X(0) = 0] = \begin{cases} \lambda_1 h + o(h), & j = 1 \\ 1 - \lambda_1 h + o(h), & j = 0 \\ 0, & o.w.. \end{cases} \tag{2.5}$$

This difference occurs due to the nonnegativity of the queue length and causes some difficulties in application of the general scheme, described in Theorem 1.3, to the large deviation principle of $M/M/1$ system, because the condition (1.12) is not satisfied. We will show the violation of the condition (1.12) in the M/M/1 queue. As we have seen in (2.4) and (2.5), the probability structure varies on the boundary of positive real line , i.e. 0. As a result, $H(x, \alpha) \neq H(x\prime, \alpha)$ if $x = 0$ and $x' > 0$. Let's see this difference carefully.

First of all,

$$A^{\epsilon} f(x) = \begin{cases} [f(x+\epsilon) - f(x)]\lambda_1(\frac{1}{\epsilon}) + [f(x-\epsilon) - f(x)]\lambda_2(\frac{1}{\epsilon}), & x > 0 \\ [f(x+\epsilon) - f(x)]\lambda_1(\frac{1}{\epsilon}), & x = 0. \end{cases}$$

Therefore, the corresponding

$$H(x,\alpha) = \begin{cases} \lambda_1(exp(\alpha)-1)+\lambda_2(exp(-\alpha)-1), & x>0 \\ \lambda_1(exp(\alpha)-1), & x=0 \end{cases} \tag{2.6}$$

and the Legendre transformation, $L(x,\beta)$, of $H(x,\alpha)$ is

$$L(x,\beta) = \begin{cases} \beta ln\frac{\beta}{\lambda_1} - \beta + \lambda_1, & x=0 \\ \beta log\frac{\beta+\sqrt{\beta^2+4\lambda_1\mu}}{2\lambda_1} - \sqrt{\beta^2+4\lambda_1\mu} + \lambda_1 + \mu, & x>0. \end{cases}$$

Then if $\lambda_2 \neq 0$, $L(x,\beta)$ in (2.6) does not satisfy the condition (1.12). The point is simply that (1.12) implies continuity in x, while (2.6) is not continuous.

Therefore, the large deviation principle for the $M/M/1$ queue length process does not follow from Wentzel's result.

Skorohod Problem Approach.

Let Y denote the birth and death process. As we saw in Section 1.5, the LDP of this process is established by the general scheme by Wentzel. The idea of our approach is to express the queue length process as a functional of the birth and death process, $X = \theta \circ Y$, and to "push the LDP through" the functional θ to obtain the rate function for the queue length process. The contraction principle (Theorem 1.4) is the main tool in pushing through. For the principle to be useful, we have to find a continuous mapping, θ , from $D = D^1[0,T]$ to $D_{(+)} = D^1_{(+)}[0,T]$ and $\theta \circ Y(t)$ has to have the same transition function as $X(t)$. The following is the answer to this question.

Theorem 2.1. *Let D_S be the set of $y(.) \in D$ such that $y(0) \geq 0$. For each $y(.) \in D_S$, there are unique pair of functions $u \in D$ and $z \in D$ satisfying*

(1) $z(t) = y(t) + u(t)$

(2) $z(t) \geq 0$

(3) *$u(.)$ is non-decreasing with $u(0) = 0$ and $u(.)$ increases only at those times t where $z(t) = 0$.*

(4) Moreover, by defining $u(.) = \sigma(y(.))$ and $z(.) = \theta(y(.))$, θ and σ are continuous mappings $D_S \to D$.

(5) Fix $\psi \in D_S$ and $\tau > 0$. Define $\psi^*(t) = \phi(\tau) + \psi(\tau + t) - \psi(\tau)$, $u^*(t) = u(\tau + t) - u(\tau)$, and $\phi^* = \phi(\tau + t)$. Then $\phi^* = \theta(\psi^*)$ and $u^* = \sigma(\psi^*)$.

This theorem is a special case of Theorem 3.5 of Chapter 3, so we will not give a separate proof here. The problem of finding $z(.)$ and $u(.)$ satisfying (1) - (4) for a given $y(.)$ is called the "Skorohod problem". We call θ the "Skorohod map".

The last step we have to go through before using the contraction principle is to verify that the process $Z = \theta \circ Y$, resulting from the mapping, θ, on the trajectories of $Y(.)$ is a version of the $M/M/1$ queue length process. With $U = \sigma \circ Y$, Z and Y are related by

$$Z(t) = Y(t) + U(t).$$

Notice that $Z(t)$ is non-negative and $U(.)$ increases only at those time t where $Z(t) = 0$. Now, we investigate the probability structure of the process $Z(t)$;

$$\begin{aligned}
&P[Z(t+h) - Z(t) = v | Z(u), 0 \le u \le t] \\
&= P[\theta(Y(.))(t+h) - \theta(Y(.))(t) = v | \theta(Y(.))(s),\ 0 \le s \le t] \\
&= P[Z(h) - Z(0) = v | Z(0)],
\end{aligned}$$

since (5) in Theorem 2.1 implies that $\theta(Y(.))(t+h) - \theta(Y(.))(t)$ depends on $Y(t+h) - Y(t)$ and $\theta(Y(.))(t)$ only and for any $s > \tau > 0$, $Y(s) - Y(\tau)$ is time homogeneous and independent of $\{Y(u) : 0 \le u \le \tau\}$. Therefore, $Z(t)$ is a time homogeneous Markov process.

When the initial state $Z(0) = i$ is a positive integer, the probability of $Z(h)$ being negative is asymptotically $o(h)$ for a sufficiently small $h > 0$, because the probability of more than one jump occurring is asymptotically $o(h)$. Since U is increasing only when $Z(t) = 0$, $U(h) = 0$ with probability of $1 - o(h)$. Therefore, $P[Z(h) - Z(0) = j | Z(0) = i]$ is the same as described in (2.4).

Suppose $Z(0) = 0$. If $Y(t)$ has one jump on [0,h] and it is a negative jump, then $Z(t) = 0$ and $U(t) = Y(t)$ on [0,h]. Since the probability of more than one jump is asymptotically o(h), the probability of $Z(h) = 0$, given $Z(0) = 0$, is asymptotically equal to the sum of the probability of $Y(h) = 0$ and the probability of $Y(h) = -1$, given $Y(0) = 0$. Therefore, $P[Z(t+h) - Z(t) = j|Z(t) = 0]$ is the same as (2.5). It follows that X and Z have a same distribution (see p.111 and p.161 Eithier and Kurtz [10]).

Since Z has the same probability distribution as X, by the contraction principle, we can assert that the set of probability measures, $\{P[X^\epsilon(.) \in .|X^\epsilon(0) = x]\}$, obeys the large deviation principle with a rate function, $J_{[0,T]}(.)$, defined by

$$J_{[0,T]}(x, \phi) = inf_{\psi|\theta(\psi)=\phi} I_{[0,T]}(x, \psi),$$

where $I_{[0,T]}$ is the rate function of the birth and death process. In particular, if A is a Borel set in $D^1[0,T]$ with respect to the uniform metric ρ_T such that $inf_{\phi\in A^o}$ $I_{[0,T]}(x,\phi) = inf_{\phi\in A}$ $I_{[0,T]}(x,\phi) = inf_{\phi\in A^c}$ $I_{[0,T]}(x,\phi)$, then

$$\begin{aligned}
&lim_{\epsilon\to 0}\epsilon log P[X^\epsilon(.) \in A|X^\epsilon(0) = x] \\
&= -inf_{\phi\in A} J(x,\phi) \\
&= -inf_{\phi\in A} inf_{\psi|\theta(\psi)=\phi} I_{[0,T]}(x,\psi)
\end{aligned}$$

For a given absolutely continuous function ϕ in $A \in D^1_{(+)}[0,T]$ with $\phi(0) = x \geq 0$,

$$J_{[0,T]}(x,\phi) = inf_{\psi|\theta(\psi)=\phi} \int_0^T L(\psi'(t))dt,$$

where $L(\beta) \equiv L(x,\beta)$ is defined in (1.18) (since L does not depend on x, we use $L(\beta)$ instead of $L(x,\beta)$).

In the next, we pass the infimum in the above expression under the integral.

Theorem 2.2. *For a given absolutely continuous function ϕ in $D^1_{(+)}[0,T]$ with $\phi(0) = x \geq 0$,*

$$inf_{\psi|\theta(\psi)=\phi} \int_0^T L(\psi'(s))ds = \int_0^t inf_{\psi(s)|\theta(\psi)=\phi} L(\psi'(s))ds.$$

We will prove this theorem in a more general setting in Theorem 4.3.

Therefore, for a given absolutely continuous function $\phi \in D^1_{(+)}[0,T]$ *with* $\phi(0) = x > 0$,

$$\begin{aligned} J(x,\phi) =& inf_{\psi|\theta(\psi)=\phi} \int_0^T L(\psi'(t))dt \\ =& \int_{T_o} inf_{\psi(t)|\theta(\psi)=\phi} L(\psi'(t))dt \ + \int_{T_1} inf_{\varphi(t)|\theta(\psi)=\phi} L(\psi'(t))dt, \end{aligned}$$

where $T_o = \{t \in [0,T] : \phi(t) = 0\}$ and $T_1 = \{t \in [0,T] : \phi(t) > 0\}$

It can be easily checked (actually it is proved in a more general setting in the proof of Theorem 4.3) that

$$J(x,\phi) = \int_{T_o} inf_{x \geq 0} L(-x)dt \ + \int_{T_1} L(\phi'(t))dt$$

$H(x,\alpha)$ is the in (2.6). Define

$$\begin{aligned} H(\alpha) =& H(x,\alpha), \text{ where } x > 0 \\ =& \lambda_1(exp(\alpha) - 1) + \lambda_2(exp(-\alpha) - 1), \\ H_1(\alpha) =& H(0,\alpha) \\ & \lambda_1(exp(-\alpha) - 1). \end{aligned}$$

$$h(x,\alpha) = \begin{cases} (H_1 \vee H)(\alpha), & x = 0 \\ H(\alpha), & x > 0. \end{cases}$$

Let $l(x,\beta)$, $L_1(\beta)$, and $L(\beta)$ be the Legendre transformation of $h(x,\alpha)$, $H_1(\alpha)$, and $H(\alpha)$, respectively;

$$L_1(\beta) = \beta ln\frac{\beta}{\lambda_1} - \beta + \lambda_1,$$

$$L(\beta) = \beta log \frac{\beta + \sqrt{\beta^2 + 4\lambda_1\lambda_2}}{2\lambda_1} - \sqrt{\beta^2 + 4\lambda_1\lambda_2} + \lambda_1 + \lambda_2),$$

and, by the property of Legendre transformation(cf. Theorem 7.6.2 in Karlin [17]),

$$l(x,\beta) = \begin{cases} \inf\{\rho_1 L(\beta_1) + \rho_2 L_1(\beta_2) : \ \beta = \rho_1\beta_1 + \rho_2\beta_2 \ ; \rho_i \geq 0, \ 1 = \rho_1 + \rho_2\}, & x = 0 \\ L(\beta), & x > 0. \end{cases}$$

Remark. *The graph of $l(0,\cdot)$ is the convex hull of the graphs of $L(\cdot)$ and $L_1(\cdot)$.*

Theorem 2.3.

$$\inf_{x \geq 0} L(-x) = l(0,0) = sup_\alpha[-h(0,\alpha)].$$

proof: As we noted in the above, the graph of l is the convex hull of the graphs of L and L_1 and $L(\beta) > 0$ except for $x = \lambda_1 - \lambda_2$ and strictly increases as $|\beta - (\lambda_1 - \lambda_2)|$ increases. Likewise $L_1(\alpha) > 0$ except for $\beta = \lambda_1$ and increases as $\alpha - \lambda_1$ increases from 0 to ∞. Since λ_1 and λ_2 are positive, $\lambda_1 > \lambda_1 - \lambda_2$. $L_1(\beta) \geq L(\beta)$ for all $0 \leq \beta \leq \lambda_1 - \lambda_2$. Therefore, $l(0,0) = \begin{cases} 0, & if\ \lambda_1 - \lambda_2 < 0 \\ L(0), & if\ \lambda_1 - \lambda_2 \geq 0. \end{cases}$ On the other hand, $\inf_{x \geq 0} L(-x) = \begin{cases} L(0), & \lambda_1 - \lambda_2 \geq 0 \\ 0, & \lambda_1 - \lambda_2 < 0 \end{cases}$ It implies that $\inf_{x\geq 0} L(-x) = l(0,0) = sup_\alpha[-h(0,\alpha)]$. ■

Therefore, for a given absolutely continuous function ϕ *with* $\phi(0) = x$,

$$\begin{aligned} J(x,\phi) &= inf_{\psi|\theta(\psi)=\phi} \int_0^T L(\psi'(t))dt \\ &= \int_{T_1} L(\phi'(t))dt + \int_{T_o} inf_{b>0} L(-b)dt \\ &= \int_0^T l(\phi(t),\phi'(t))dt. \end{aligned}$$

In this section, we have successfully shown how the LDP of $M/M/1$ process can be obtained by the Skorohod problem approach. In the following chapters, we will generalize this approach to the class of queueing networks described in Section 2.1.

Chapter 3. Large Deviation Principle for Queueing Systems

In this chapter, we want to develop the Skorohod problem approach to find the LDP of the general OCOS type Jackson network (we call it OCOS system). For this purpose, we define the "potential process" of the system, which corresponds to the birth and death process in the case of M/M/1 system. The LDP of the potential process is easily obtained by Wentzel's result. We formulate the Skorohod problem for an OCOS system and find 'Skorohod map' which maps the potential process to the queue length process. By using it, we find the LDP of the system by applying the contraction principle.

3.1 The Potential Process and Its LDP.

The discontinuity (cf. (1.12)) in the probability structure of a queueing system is mainly caused by the nonnegativity condition on the queue length. If there is no customer in a station, then the station is in an idle state and no service is offered. This can be thought of as a nonnegativity constraint on the queue length process, which imposes a change in the probabilistic behavior of the server when its queue becomes empty. This is the main obstacle in finding the LDP of the queueing system through Wentzel's results. In order to avoid this nonnegativity constraint, we consider the potential process which is described as follows. Let's consider a system in which there is only one server at each station and servers provide units of services. They are not concerned whether there are customers in their station or not. The time which is needed for the server in the station i in generating a unit of service follows the exponential distribution with parameter $\lambda_{v(i)}$. If they provide a service unit at a time when their queue is empty, then we consider the "queue length" to change from 0 to -1. Similarly if the queue length is already negative we simply decrement it by

1 again at the service time. The rest of the topological structure and probabilistic assumptions are defined same as the one for OCOS system. The variable we observe at time t at each station is the number of arrivals up to time t to the station minus the number of the service units generated in the station up to time t. Let $Y_i(t)$ be the observation in i^{th} station at time t and $Y(t)$ be the N-dimensional process whose i^{th} element is $Y_i(t)$. We call Y "potential process".

Large Deviation Principle of the Potential Process.

We begin with any probability space (Ω, σ, P) on which the potential process $Y = \{Y(t)|0 \leq t < \infty\}$ is defined. Let $\sigma_t = \sigma(Y(s); 0 \leq s \leq t)$ for $t \geq 0$.

Consider the family of stochastic processes $\{Y^\epsilon(t)\} \equiv \{\epsilon Y(\frac{t}{\epsilon})\}$. We want to specify the LDP of $\{P \circ Y^{\epsilon\ -1}\}$. By applying Theorem 1.3, we can find the rate function of the LDP.

First, we find the small time increment conditional probability with an initial state x;

$$P[Y(h) - Y(0) = v|Y(0) = x] = \begin{cases} \lambda_v h + o(h), & nonzero\ v \in V \\ 1 - \sum_{v \in V} \lambda_v h + o(h), & v = o \\ o(h), & o.w.. \end{cases} \tag{3.1}$$

Notice that this probability does not depend on the initial state x. Based on this generator, we find the function H which is given by (1.8) in Chapter 1.

$$H(x, \alpha) = \sum_{v \in\ V} \lambda_v(exp(< \alpha,\ v >) - 1),\ \forall x \in R^N. \tag{3.2}$$

Since $H(x, \alpha)$ does not depend on the state variable, x, we will simply write $H(\alpha)$ for $H(x, \alpha)$ throughout the following. It can be noticed that $H(\alpha)$ is strictly convex and analytic on R^N and $H(0) = 0$. Since each station is assumed to have exactly one channel for outflow and at least one channel for inflow, for each i, there exist exactly one $v \in V$ with $v_i = -1$ and at least one $v \in V$ with $v_i = 1$. Therefore, as $\|\alpha\|$ goes to ∞,

$$H(\alpha) \text{ goes to } \infty \tag{3.3}$$

Its Legendre transformation is defined by

$$L(\beta) = sup_{\alpha \in R^N}[< \alpha, \beta > -H(\alpha)],\ \beta \in R^N. \tag{3.4}$$

Since $H(0) = 0$, $L(\beta)$ is nonnegative, for all β in R^N(cf. (1.3) Chapter 1). Since $H(\alpha)$ is convex and analytic, $L(\beta)$ is strictly convex and analytic on R^N(cf. (1.5) in Chapter 1). And as we said in Section 1.1, since $H(\alpha)$ is bounded in any bounded domain, $L(\beta)$ goes to ∞, as $\|\beta\|$ goes to ∞. Since H and L do not depend on the state variable, we can easily check the conditions (1.10) - (1.12) in Chapter 2. Therefore, Theorem 1.3 applies to $Y^\epsilon(t) = \epsilon Y(t/\epsilon)$, giving an LDP with rate function

$$I_{[0,T]}(x, \phi) = \begin{cases} \int_0^T L(\phi'(t))dt, & \phi \textit{ is a absolutely continuous function}, \\ & \textit{its integral is convergent, and } \phi(0) = x \\ \infty, & o.w.. \end{cases} \tag{3.5}$$

3.2 Skorohod Problem in OCOS System.

Our objective in this section is to exhibit the queue length process of an OCOS system as a continuous functional of the potential process Y: $Z \equiv \theta \circ Y$. Z should be a time homogeneous N dimensional Markov jump process on the nonnegative orthant $R^N_{(+)} = \{x \in R^N : \text{all} x_i \geq 0\}$. . Second, Z should behave on the interior of $R^N_{(+)}$ like the potential process Y. Third, Z reflects everywhere on the boundary of $R^N_{(+)}$ so that Z has the same distribution as the queue length process of the OCOS system X. In particular, the direction of reflection on i^{th} boundary surface is fixed and has the opposite direction to the column vector v whose i^{th} element is -1. (Recall that there is only one such v in V and from now on, we write it $v(i)$.) But on "edges" there are multiple reflective directions. What we want to do is formulate an appropriate analogue of the Skorohod problem of Theorem 2.1 above.

We can put the properties above into precise mathematical terms as follows; we seek a pair of N-dimensional processes $Z \equiv \{Z(t); t \geq 0\}$ and $U \equiv \{U(t); t \geq 0\}$ which jointly satisfy the following conditions;

$$Z(t) = Y(t) - AU(t) \tag{3.6}$$

where Y is the potential process with $Y(0) \in R^N_{(+)}$ and A is the $N \times N$ matrix whose i^{th} column A_i is the $v \in V$ whose i^{th} element is -1, $v(i)$.

$$Z(t) \in R^N_{(+)}, \forall t \in [0, T]. \tag{3.7}$$

and, for each i (and almost all ω),

$$U_i(\omega, \cdot) \in \; D^N[0, \infty) \text{ is nondecreasing with } U_i(\omega, 0) = 0. \tag{3.8}$$

and

$$U_i(\omega, \cdot) \text{ increases only at those times } t \text{ when } Z_i(\omega, t) = 0. \tag{3.9}$$

The rest of this section is devoted to showing that there is a continuous mapping θ from $D^N[0, T]$ to $D^N_{(+)}[0, T]$ which maps the trajectories of Y into those of Z satisfying (3.6)-(3.9) trajectorywise:

$$\theta(Y(\omega, \cdot)) = Y(\omega, \cdot) - AU(\omega, \cdot). \tag{3.10}$$

The general formulation of the Skorohod problem given in Definition 3.1 below, as well as Theorem 3.2 giving sufficient conditions for its well-posedness are taken from Dupuis and Ishii [8]. In Theorem 3.6, we formulate its implications in the special context appropriate for our OCOS queueing systems.

Let $R_i = \{x \in R^N : x_i \geq 0\}$ and $R^N_{(+)}$ be the positive orthant of R^N. We may, therefore, define $R^N_{(+)} = \cap_i^N R_i$. For $\eta \in D^N[0,T]$, let $|\eta|(T)$ denote the total variation of η on $[0,T]$. For each i, let n_i be the inward normal vector to $R^N_{(+)}$ on R_i. Suppose that, for each j, d_j is a unit vector which satisfies

$$\langle d_j, n_j \rangle > 0. \tag{3.11}$$

For each $x \in \partial R^N_{(+)}$,

$$d(x) = \{\gamma = \sum_{i \in B(x)} \alpha_i d_i ; \alpha_i \geq 0, \|\gamma\| = 1\}$$

and

$$n(x) = \{\gamma = \sum_{i \in B(x)} \alpha_i n_i ; \alpha_i \geq 0, \|\gamma\| = 1\}.$$

We consider $d(x)$ to be the allowable set of "directions of reflection"at x.

Definition 3.1. *(Skorohod Problem(SP)) Let $\psi \in D^N[0,T]$ with $\psi(0) \in R^N_{(+)}$ be given. Then $(\phi, \psi, \eta\ ) \in [D^n[0,T]]^3$ solves the Skorohod problem (SP) (or (ϕ, η) solves the SP for ψ) with respect to $d(\cdot)$ if*

(1) $\phi = \psi + \eta$, $\phi(0) = \psi(0)$,

(2) $\phi(t) \in R^N_{(+)}$ *for* $t \in [0,T]$,

(3) $|\eta|(T) < \infty$,

(4) $|\eta|(t) = \int_{(0,t]} I_{\phi(s) \in \partial R^N_{(+)}} d|\eta|(s)$,

(5) there exists measurable $\gamma : [0,T] \to R^N$ such that $\gamma(s) \in d(\phi(s))(d|\eta|$ a.s.) and $\eta(t) = \int_{(0,t]} \gamma(s) d|\eta|(s)$.

Hence ϕ never leaves $R^N_{(+)}$ and η changes only when $\phi \in \partial R^N_{(+)}$, in which case the change points in one of the directions $d(\phi(\cdot))$.

The following theorem from [8] tells us conditions under which the SP has a unique solution.

Theorem 3.2. *Assume that there exist positive constants a_i such that*

$$a_i \langle n_i, d_i \rangle > \sum_{j \neq i} a_j |\langle n_i, d_j \rangle| \tag{1}$$

for all i. Define the $N \times N$ matrix D by $d_i = D n_i, 1 \leq i \leq N$.
Assume that there exist $c > 0$ such that

$$\langle n, Dn \rangle \geq c > 0, \quad \textit{for all } n \in n(x), \textit{ all } x \in \partial R^N_{(+)}, \tag{2}$$

and for each $x \in \partial R^N_{(+)}$, there is $n \in n(x)$ such that for all $d \in d(x)$,

$$\langle d, n \rangle > 0. \tag{3}$$

With these assumptions, the SP has a unique solution for every $\psi \in D^N[0,T]$ with $\psi(0) \in R^N_{(+)}$. Moreover, let (ϕ_1, ψ_1, η_1) and (ϕ_2, ψ_2, η_2) be the corresponding solutions of the SP. Then there exists $K < \infty$ (which is independent of T) such that

$$\rho_T(\eta_1, \eta_2) \leq K \rho_T(\psi_1, \psi_2),$$

$$\rho_T(\phi_1, \phi_2) \leq K \rho_T(\psi_1, \psi_2),$$

when ρ_T is the uniform metric in $D^N[0,T]$.

This is essentially Theorem 3.3 and 3.4 of [8]. Their Assumption 2.1 is implied by our (1) and their Theorem 2.1. Their Assumption 3.1 is implied by our (2) and their

Theorem 3.1. (3) is a sufficient condition for Assumption 3.2 and (1) is a sufficient condition for (3.11). Therefore, the conclusion of the theorem follows from the results in [8].

Let $Q \equiv I + A$. Then Q is a lower triangular matrix with zero's on the diagonal. Every element in Q is nonnegative and Q has $\{0\}$ as its spectrum. Thus there exists a diagonal matrix Λ, having positive diagonal elements, such that the maximal row sum of the nonnegative matrix $Q^* = \Lambda^{-1}Q\Lambda$ is strictly less than 1 (cf., Lemma 3 in Veinott [27]). Q^* is also a lower triangular matrix and the locations of nonzero elements of Q^* are same as those of Q. (Let M_i be the i^{th} column of a matrix M.)

Defining $d_i = \frac{1}{\|(I-Q^*)_i\|}(I - Q^*)_i$, we check in the following if the assumed conditions (1)-(3) in Theorem 3.2 are satisfied with these d_i's.

Condition (1): Without loss of generality, instead of the condition (1), we check the existence of positive constant a_i's which satisfy

$$a_i\langle n_i, (I - Q^*)_i\rangle > \sum_{j \neq i} a_j |\langle n_i, (I - Q^*)_j\rangle|, \ \forall i \in \{1, \cdots, N\}.$$

When $i \neq j$, $\langle n_i, (I - Q^*)_j\rangle \leq= 0$, since elements of Q^* are positive and the location of off-diagonal elements are same as that of Q. Hence the condition above is equivalent to

$$a_i\langle n_i, (I - Q^*)_i\rangle > \sum_{j \neq i} a_j \langle n_i, -(I - Q^*)_j\rangle, \ \forall i \in \{1, \cdots, N\}.$$

Then, since the elements on the diagonal of the matrix $I - Q^*$ are equal to 1, this condition is equivalent to

$$(I - Q^*)a > 0.$$

where $a = (a_1, \cdots, a_N)^T$. Since the maximal row sum of the positive matrix Q^* is less than 1, the above condition is satisfied with any vector a which has the following property;

$$a_1 > 0 \ and \ a_i > \sum_{1 \le j < i} a_j.$$

Condition (2): $I - Q^*$ is a lower triangular matrix whose diagonal terms are equal to 1 and the maximal row sum of Q^* is less than 1. Therefore, the minimal row sum of $I - Q^*$ is greater than 0. Since the location of nonzero elements of Q are same as those of Q^* and Q has only one off-diagonal term in each column, Q^* has only one nonzero element in each column. Also each element of Q^* is less than 1. Therefore, the maximal column sum of Q^* is less than 1 and, therefore, the minimal row sum of $(I - Q^*)^T$ is greater that 0. It follows that $I - Q^*$ is a positive definite matrix (cf., Golub [14], p.7). The set of unit vectors, $n(0)$, is a closed subset of $R^N_{(+)}$ which does not contain 0. Therefore, there is a positive constant c such that $\alpha^T(I - Q^*)\alpha > c$, for all $\alpha \in n(0)$. For any $x \in \partial R^N_{(+)}$, $n(x) \subset n(0)$ and, therefore, $\alpha^T(I - Q^*)\alpha > c$, for all $\alpha \in n(x)$. This is a sufficient condition for Assumption 3.1 in [8]. (In fact, if there is a positive diagonal matrix Σ such that $D\Sigma$ satisfy the condition (2), the proof of Theorem 3.1 in [8] is applicable.)

Condition (3): Let $x \in \partial R^N_{(+)}$. Choose a normal vector $n^* = \sum_{i \in B(x)} \beta_i n_i$ with $\beta_i > 0$ for all i and $\beta_i > \beta_j > 0$ for $i < j$ in $B(x)$. If $d \in d(x)$, then $d = \sum_{i \in B(x)} \alpha_i d_i$. Since d is a unit vector, there is at least one $j \in B(x)$ such that $\alpha_j > 0$. Therefore, for all $d \in d(x)$, $\langle n^*, d \rangle > 0$.

Therefore, we can claim the following corollary.

Corollary 3.3. *If $d_i = \frac{1}{\|(I-Q^*)_i\|}((I-Q^*)_i)$ is the direction of reflection on each R_i, the SP has a unique solution for any function ψ in $D^N[0,T]$. Moreover, let (ϕ_1, ψ_1, η_1) and (ϕ_2, ψ_2, η_2) be the corresponding solutions of the SP Then there exists $K < \infty$ (which is independent of T) such that*

$$\rho_T(\eta_1, \eta_2) \le K \rho_T(\psi_1, \psi_2),$$

$$\rho_T(\phi_1, \phi_2) \le K \rho_T(\psi_1, \psi_2). \tag{3.12}$$

In the following, we reformulate the SP above into an equivalent form, which we will refer to as SP*. We do this to express the directions of reflection more explicitly and explain the Skorohod problem in a more intuitive way. We also seek notational simplicity for the next section. Let e_i be the unit vector in R^N whose i^{th} element is 1.

Lemma 3.5. *For a given $\psi(.) \in D^N[0,T]$ for $\psi(0) \in R^N_{(+)}$, there is a unique pair of functions u, $\phi \in D^N_{(+)}[0,T]$ satisfying the following (which we refer to as SP*):*

(1) $\phi(t) = \psi(t) + (I - Q^*)u(t)$

(2) $\phi_j(t) \ge 0, j = 1, 2, ..., N$

(3) $u_j(.)$ is non-decreasing with $u_j(0) = 0$ and $u_j(.)$ increases only at those times t where $\phi_j(t) = 0$.

Proof: Let (ϕ, η) be the solution of the SP with the given ψ. Let $u \equiv (I - Q^*)^{-1}\eta$. (1) and (2) are trivially satisfied and $\eta(0) = 0$ implies $u(0) = 0$. Therefore, it is sufficient to prove that u_i is nondecreasing and increases only at those times t where $\phi_i(t) = 0$. By (5) in definition 3.1,

$$(I - Q)^{-1}\eta(t) = \int_{(0,t]} (I - Q)^{-1}\gamma(s) d|\eta|(s).$$

By (4) in definition 3.1, $|\eta|(t)$ is increasing only when $\phi(t) \in \partial R^N_{(+)}$. And $\gamma(s) = \sum_{i\in B(\phi(s))} \alpha_i d_i$ for some positive α_i's. $(I - Q^*)^{-1}\gamma(s) = (I - Q^*)^{-1} \sum_{i\in B(\phi(s))} \alpha_i d_i$ $= \sum_{i\in B(\phi(s))} \alpha_i (I - Q^*)^{-1} d_i$. If $v = (I - Q^*)^{-1} d_i$, then $(I - Q^*)v = d_i = \frac{(I-Q^*)_i}{\|(I-Q^*)_i\|}$ and, therefore, $v = \frac{1}{\|(I-Q^*)_i\|} e_i$. It tells that $(I - Q^*)^{-1} d_i$ is a column vector whose elements are all zero's except the i^{th} element which is positive. Therefore, $u_i(t) = ((I - Q^*)^{-1}\eta(t))_i$ increases only when $\phi_i(t) = 0$ and $u_i(t)$ is nondecreasing. ∎

However, the set of natural directions of reflection on boundary surfaces is not the set of $(I - Q^*)_i'$s, but the set of $-A_i$'s. In fact, we are looking for the solution of the SP^* with $d_i = -\frac{v(i)}{\|v(i)\|} = \frac{(I-Q)_i}{\|(I-Q)_i\|}$, not with $d_i = \frac{(I-Q^*)_i}{\|(I-Q^*)_i\|}$. But, as we saw, $I - Q$ can be decomposed into $\Lambda(I - Q^*)\Lambda^{-1}$ and Λ is a positive diagonal matrix. Therefore, we can claim the following theorem.

Theorem 3.6. *For a given $\psi(.) \in D^N[0,T]$ with $\psi(0) \in R^N_{(+)}$, there is a unique pair of functions (ϕ, u), satisfying the following (which we refer to as SP^{**}):*

(1) $\phi(t) = \psi(t) - Au(t)$

(2) $\phi_j(t) \geq 0,\ j = 1, 2, ..., N$

(3) *$u_j(.)$ is non-decreasing with $u_j(0) = 0$ and $u_j(.)$ increases only at those times t where $\phi_j(t) = 0$.*

Proof: Let $\psi^*(t) = \Lambda^{-1}\psi(t)$. Since Λ is a diagonal matrix with positive diagonal elements and $\psi(0) \in R^N_{(+)}$, $\psi^*(0) \in R^N_{(+)}$. Let ϕ^* and u^* be the unique solution of the SP^* for the given ψ^* with the matrix $I - Q^*$, i.e. $\phi^*(t) = \psi^*(t) + (I - Q^*)u^*(t)$. Then, the uniqueness of the SP^* solution and nonsingularity of Λ tells that $(\Lambda\phi^*, \Lambda u^*)$ is the unique solution to the SP^{**};

Let $(\phi, u) = (\Lambda\phi^*, \Lambda u^*)$. Then, first,

$$\begin{aligned}\phi(t) &= \psi(t) + \Lambda(I - Q^*)\Lambda^{-1}u(t) \\ &= \psi(t) - Au(t).\end{aligned}$$

Second, since Λ is a nonnegative matrix and $\phi_j^* \geq 0$, for all j, $[\Lambda\phi^*]_j \geq 0$, for all j. Since u_j^* is nondecreasing and increases only at those times t where $\phi_j^*(t) = 0$ and Λ is a diagonal matrix with positive diagonal elements, $(\Lambda u^*)_j$ is nondecreasing and increases only at those times t where $(\Lambda\phi^*(t))_j = 0$. If there is another solution of the SP^{**} with ψ, there must be another solution of the SP^* with $\Lambda^{-1}\psi$. It contradicts the uniqueness of the solution of the SP^*. Therefore, there is only one solution of the SP^{**}. ■

From now on, we write SP as SP^{**}. The reason that we define and solve the SP is to find a continuous mapping which maps the potential process to a version of the queue length process of OCOS system and obtain the LDP of the queue length process by the contraction principle. So, as the next step, we specify the mapping explicitly and investigate its properties carefully.

Let $D_S^N = \{\psi \in D_{(+)}^N[0,T] \text{ with } \psi(0) \in R_{(+)}^N\}$. Let (ϕ, u) is the solution of SP for $\psi \in D_S^N$ We set $u(\cdot) = \sigma(\psi(\cdot))$ and $\phi(\cdot) = \theta(\psi(\cdot))$.

Lemma 3.7.

θ and σ are continuous mappings $D^N[0,T] \to D^N[0,T]$ with respect to ρ_T.

(1) The restriction of ψ and ϕ to $[0,t]$ depends only on the restriction of x to $[0,t]$.

(2) Fix $\psi \in D_S^N$ and $\tau > 0$. Let $\phi = \theta(\psi)$ and $u = \sigma(\psi)$ as above. Define $\psi^*(t) = \phi(\tau) + \psi(\tau + t) - \psi(\tau)$, $u^*(t) = u(\tau + t) - u(\tau)$, and $\phi^*(t) = \phi(\tau + t)$.

Then $\phi^* = \theta(\psi^*)$ and $u^* = \sigma(\psi^*)$.

Proof: (3.12) says $\rho_T(\theta(\psi_1), \theta(\psi_2)) \leq K\rho_T(\psi_1, \psi_2)$, which is much stronger than the continuity. (1) is trivially satisfied by the definition of solution of the SP. To prove (2), given the uniqueness statement of the theorem, one need only verify that (ϕ^*, ψ^*, u) solves the SP. For (1),

$$\begin{aligned}\psi^*(t) - Au^*(t) =& \phi(\tau) + \psi(\tau + t) - \psi(\tau) - Au(\tau + t) + Au(\tau)\\ =& \psi(\tau + t) - Au(\tau + t)\\ =& \phi(\tau + t) = \phi^*(t).\end{aligned}$$

(2) and (3) in Theorem 3.6 are trivially satisfied. ■

We name θ the "Skorohod map". Based on this map, we construct a process Z on (Ω, σ, P) on which the potential process Y is defined. To begin with, let $Z(\omega)(t) = \theta(Y(\omega))(t)$, $U(\omega) = \sigma(Y(\omega))(t)$ on Ω. In Lemma 3.7, we investigate the probability structure of the process Z.

Lemma 3.8.

(1) Both Z and U are measurable w.r.t. σ_t for each $t \geq 0$.

(2) The processes Z and U satisfy the following almost surely;

1. $Z(t) \in R^N_{(+)}$, $\forall t \geq 0$.

2. $U_i(\omega, t)$ is nondecreasing with $U_i(\omega, 0) = 0$.

3. $U_i(\omega)$ increases only at those time t where $Z_i(\omega, t) = 0$, $\forall i = 1, 2, ..., N$.

(3) Z is a time homogeneous Markov process on (Ω, σ, P).

Proof: Y is measurable with respect to σ_t for each $t \geq 0$ and θ is continuous. It implies (1). (2) follows directly from Theorem 3.6. By Lemma 3.7, we notice that $Z(t + h) - Z(t)$ is solely dependent upon $Z(t)$ and $Y(t + h) - Y(t)$. Since for any

$h > 0$, $Y(t+h) - Y(t)$ is independent of $\{Y(s), 0 \leq s \leq t\}$ and time homogeneous. Z is a time homogeneous Markov process. ■

Next, we want to show that Z is a version of the queue length process X of OCOS system. First, we calculate the small time increment conditional probability at an initial state. Since Z is a time homogeneous Markov process, for any Borel measurable set $\mathcal{O}$ in R^N,

$$
\begin{aligned}
&P[Z(t+h) - Z(t) \in \mathcal{O} | Z(u), 0 \leq u \leq t] \\
&= P[Z(t+h) - Z(t) \in \mathcal{O} | Z(t)] \\
&= P[Z(h) - Z(0) \in \mathcal{O} | Z(0)].
\end{aligned}
$$

It follows that the above probability is equal to

$$
\begin{aligned}
&P[\theta(Y)(h) - \theta(Y)(0) \in \mathcal{O} | \theta(Y)(0)] \\
=&P[\theta(Y)(h) - Y(0) \in \mathcal{O} | Y(0)] \\
=&P[Y(h) - AU(h) - Y(0) \in \mathcal{O} | Y(0)]
\end{aligned}
$$

The potential process of OCOS system can be considered as a linear function of joint Poisson processes. This means that the probability of the process Y having more than one jump in a small amount of time h is equal to $o(h)$.
In particular, as h decreases to 0,

$$
\sum_{v \in V \cup \{0\}} P[Y(h) - Y(0) = v | Y(0) = x] = 1 + o(h).
$$

Recall that

$$
P[Y(h) - Y(0) = v | Y(0) = x] = \begin{cases} \lambda_v h + o(h), & \textit{nonzero } v \in V \\ 1 - \sum_{v \in V} \lambda_v h + o(h), & v = 0 \\ o(h), & o.w.. \end{cases} \tag{3.13}
$$

Remark. *Each $v \in V$ represents one type of jump in the process Y. Moreover, $Y(\omega,t)$ is in $\mathbb{Z}^N$, almost surely, and given $Y(t) \neq Y(t^-)$, $Y(t)-Y(t^-) \in V$, almost surely.*

Suppose $Y(\omega,\cdot)$ is an N dimensional step function on $[0,T]$ with $Y(\omega,0) \in \mathbb{Z}^N_{(+)}$ and has only one discontinuity at τ on $[0,h]$ when $h>0$. Let $Y(\omega,\tau)-Y(\omega,\tau^-)=v$ for some $v \in V$. From (1) in Theorem 3.6,

$$Y(\omega,h)-Y(\omega,0)=Z(\omega,h)-Z(\omega,0)-A(U(\omega,h)).$$

Therefore,

$$v=Z(\omega,h)-Z(\omega,0)-A(U(\omega,h)).$$

If $v_i \geq 0$ for all $i \in B(Z(\omega,0))$, $Z(\omega,0)+v \geq 0$. By letting $Z(\omega,\cdot) \equiv Y(\omega,\cdot)$ on $[0,h]$, $(Z(\omega,\cdot),0)$ is the unique solution of SP for $Y(\omega,\cdot)$ on $[0,h]$. Therefore,

$$v=Z(\omega,h)-Z(\omega,0).$$

If $v=v(i)$ for some $i \in B(Z(\omega,0))$, $Z_i(\omega,0)+v_i=-1$. By letting $Z(\omega,\cdot)=Z(\omega,0)$ and $U(\omega,t)=\begin{cases} 0 & \textit{on } [0,\tau) \\ e_i & \textit{on } [\tau,h], \end{cases}$, $(Z(\omega),U(\omega))$ is the unique solution of SP for $Y(\omega,\cdot)$ on $[0,h]$. In particular,

$$AU(\omega,h)=v \textit{ and } Z(\omega,h)-Z(\omega,0)=0.$$

It follows that

$$Y(\omega,h)-Y(\omega,0)=v=\begin{cases} Z(\omega,h)-Z(\omega,0) & \textit{when } v \notin V_{B(Y(\omega,0))} \\ AU(\omega,h) & \textit{when } v \in V_{B(Y(\omega,0))}, \end{cases}$$

where $V_K=\{v(i) : i \in K\}$.

For $i \in B(x)$ where $x \in \mathbb{Z}^N_{(+)}$, the event

$$\{\omega : Y(\omega, h) - Y(\omega, 0) = v(i), Y(\omega, \cdot) \text{ has only one jump on } [0, h] | Y(\omega, 0) = x\}$$

is a subset of

$$\begin{aligned}
&\{\omega : Y(\omega, h) - AU(\omega, h) - Y(\omega, 0) = 0, \\
&\qquad\qquad Y(\omega, \cdot) \text{ has only one jump on } [0, h] | Y(0) = x\} \\
=&\{\omega : Z(\omega, h) - Z(\omega, 0) = 0, Y(\omega, \cdot) \text{ has only one jump on } [0, h] | Z(0) = x\}.
\end{aligned}$$

For v with $v_i \geq 0$ for all $i \in B(x)$ where $x \in \mathbb{Z}^N_{(+)}$, the event

$$\{\omega : Y(\omega, h) - Y(\omega, 0) = v, Y(\omega, \cdot) \text{ has only one jump on } [0, h] | Y(\omega, 0) = x\}$$

is equal to

$$\{\omega : Z(\omega, h) - Z(\omega, 0) = v, Y(\omega, \cdot) \text{ has only one jump on } [0, h] | Z(0) = x\}.$$

Therefore,

$$P[Z(h) - Z(0) = v | Z(0) = x] = \begin{cases} \lambda_v(x)h + o(h), & \text{nonzero } v \in V \\ 1 - \sum_{v \in V} \lambda_v(x)h + o(h), & v = 0 \\ o(h), & o.w.. \end{cases}$$

It implies that the processes Z and the queue length process X have the same small time increment conditional probability on each initial condition x. Since they are Markov jump processes which have the same small time increment conditional probability, they have the same infinitesimal generator. It follows that they have the same distribution.(cf. p.111 and p. 161 in Ethier and Kurtz [10].)

3.3 The Rate Function of the Queue Length Process.

The goal of this chapter is to construct the LDP of the process X. However, as we found in the last section, the process Z has the same distribution as the process X. Then they follow the same LDP. It is simply because LDP of a Markov process is determined by its distribution. Therefore, we construct the LDP of Z.

Consider the rescaled processes X^ϵ, Z^ϵ;

$$X^\epsilon = \epsilon X(\frac{t}{\epsilon}), \quad Z^\epsilon = \epsilon Z(\frac{t}{\epsilon}).$$

For Borel sets $G \in D^N_{(+)}[0,T]$, let

$$\begin{aligned} Q^\epsilon(G) &\equiv Pr[Z^\epsilon \in G | Z^\epsilon(0) = x] \\ &= Pr[Y^\epsilon \in \theta^{-1}(G) | Y^\epsilon(0) = x] \\ &= Pr[\theta(Y^\epsilon) \in G | Y^\epsilon(0) = x] \\ &\equiv P^\epsilon[\theta^{-1}(G)]. \end{aligned}$$

Therefore, by the contraction principle (Theorem 1.4), the set of probability measures, Q^ϵ, follows the large deviation principle with the rate function, $J_{[0,T]}(\cdot)$, defined by

$$J_{[0,T]}(x, \phi) = \inf_{\psi | \theta(\psi) = \phi} I_{[0,T]}(x, \psi), \tag{3.14}$$

where $I_{[0,T]}$ is the rate function for LDP of the potential process in (3.5). Since

$$Pr[Z^\epsilon \in G | Z^\epsilon(0) = x] = Pr[X^\epsilon \in G | X^\epsilon(0) = x],$$

we can assert that $J_{[0,T]}(x, \cdot)$ is the rate function of the queue length process of this system. Using the expression in (3.5) for $I_{[0,T]}$, (3.14) provides the following theorem.

Theorem 3.9. *If ϕ is absolutely continuous on $[0,T]$ with $\phi(0) = x$ and the following is defined,*

$$\begin{aligned} J_{[0,T]}(x, \phi) &= \inf_{\psi | \theta(\psi) = \phi} \int_0^T L(\psi'(t)) dt. \\ &= \inf_{u | \theta(\psi) = \phi,\ \sigma(\psi) = u} \int_0^T L(\phi'(t) + A u'(t))) dt. \end{aligned}$$

Otherwise, it is ∞.

The latter expression is because $\theta(\psi) = \phi$ and $\sigma(\psi) = u$ implies that $\phi = \psi - Au$.

As we have seen in the above, the queue length process X of OCOS system can be understood in the framework of LDP and we can obtain the rate function of $\{P^\epsilon = P \circ (X^\epsilon)^{-1}\}$ in the form of infimum of the rate function of potential process over a corresponding set of trajectories of the potential process.

Chapter 4. The Rate Function of Queueing Systems

The goal of this chapter is to obtain a closed form expression for (3.14). Even though we use a completely different approach, we will find that our closed form expression is same as that discovered by Dupuis in [9]. In Section 4.2 we will explain why the Skorohod problem approach that we presented in this thesis does not work in a more general class of Jackson networks.

4.1 The Rate Function in a Closed Form.

To find the closed form of (3.14), we pass the infimum under the integral. This will be done in Theorem 4.3. In preparation for this we develop some lemmas.We will frequently use Lemma A.1 - A.4 which can be found in Appendix.

Let R_K^N=$\{x \in R^N | x_i = 0, \forall i \in K\}$, where K is a subset of $\{1, ..., N\}$. And $R_K^{N(+)}$ =$\{x \in R_K^N | x_i \geq 0,\ i \in K^c\}$,

Lemma 4.1. *Suppose $f(.)$ is a strictly convex continuous function on R^N and as $\|x\| \to \infty$, $f(x) \to \infty$. Define $g(x) = \inf_{c \in R_{K^c}^{N(+)}} f(x+c)$, for $x \in R_K^N$ Then , for each $x \in R_K^N$, there exists a unique $c(x) \in R_{K^c}^{N(+)}$ s.t. $g(x) = f(x+c(x))$. Moreover, c is a continuous mapping from R_K^N to $R_{K^c}^{N(+)}$.*

Proof: Let C is a closed convex set of R^N. Since as $\|x\| \to \infty$, $f(x) \to \infty$, if there is a optimal point x^* $s.t.$ $\inf_{x \in C} f(x) = f(x^*)$, then $\|x^*\| < \infty$.

Suppose that there are two optimal points x_1 and x_2 such that $\inf_{x \in C} f(x) = f(x_1) = f(x_2)$. Then x_1 and x_2 are in C. Since f is strictly convex, for $0 \leq t \leq 1$,

$$f(tx_1 + (1-t)x_2) < tf(x_1) + (1-t)f(x_2) = f(x_1)$$

and $tx_1 + (1-t)x_2 \in C$. This contradicts the fact that $\inf_C f(x) = f(x_1)$. Therefore, $f(.)$ has a unique minimum on C. In consequence, $g(x) = f(x - c(x))$, for some $c(x) \in R^{N(+)}_{K^c}$. Therefore the mapping c is well-defined.

Let $\{x_n\}$ be a sequence which converges to x and $x_n, x \in R^N_K$. From Lemma A.4, we know that g is continuous. By continuity, $g(x_n) \to g(x)$, i.e. $f(x_n - c(x_n)) \to f(x - c(x))$ and $\{f(x_n - c(x_n))\}$ is a bounded sequence. This implies that $\{c(x_n)\}$ is a bounded sequence such that $f(x_n - c(x_n)) \to f(x - c(x))$. Suppose $\{x_n - c(x_n)\}$ does not converge to $x - c(x)$. Then there exists a subsequence $\{x_{n_k} - c(x_{n_k})\}$ converging to $y (\in R^N) \neq x - c(x)$. However, $f(x_n - c(x_n)) \to f(x - c(x))$ and also $f(x_{n_k} - c(x_{n_k})) \to f(y)$ as $k \to \infty$ and $f(y) = f(x - c(x))$. Since $x_n \to x$ and $x_{n_k} - c(x_{n_k})\}$ converges to y, $c(x_{n_k}) \to (x - y) \in R^{N(+)}_{K^c}$. Therefore, the fact that $f(x - c(x)) = f(x - (x - y))$ and $x - y \neq c(x)$ contradicts the uniqueness of the local minimum. Hence, $x_n - c(x_n)$ converges to $x - c(x)$. So, $c(x_n)$ converges to $c(x)$. This shows that c is continuous. ■

Let K be a subset of $\mathbb{Z}_N$. Let $\xi^K_i = \langle \alpha, v(i) \rangle$ for $i \in K$ and $\xi^K_i = -\alpha_i$ for $i \in K^c$. Define the matrix B^K such that its i^{th} row is $v(i)$ for $i \in K$ and its j^{th} row is $-e_j$ for $j \in K^c$. The change of variable $B^K \alpha = \xi^K$, will be convenient for some of our arguments below. B^K is a nonsingular $N \times N$ matrix. We will use $H^K(\cdot)$ to indicate $H(\cdot)$ with respect to ξ^K.

$$
\begin{aligned}
H^K(\xi^K) &= H((B^K)^{-1}\xi^K) \\
&= \sum_{v \in V} \lambda_v (exp(\langle (B^K)^{-1}\xi^K, v \rangle) - 1) \\
&= \sum_{v \in V} \lambda_v (exp(\langle \alpha, v \rangle) - 1) \\
&= H(\alpha).
\end{aligned}
$$

As $\|\xi^K\| \to \infty$, $\|\alpha\| \to \infty$. Therefore, as $\|\xi^K\| \to \infty$, $H^K \to \infty$. Clearly, H^K is strictly convex and smooth, since H has these properties. Let L^K denote the Legendre transformation of H^K.

Lemma 4.2. *Let $B(x) = K$. For any $\beta \in R^N_{K^c}$,*

$$L(x + A\beta) = L^K(-x + \beta).$$

Proof: Let $\beta \in R^N_{K^c}$.

$$\begin{aligned}
&L(x + A\beta) \\
=&sup_{\alpha\in R^N}[\langle x + A\beta, \alpha\rangle - H(\alpha)], \\
=&sup_{\alpha\in R^N}[\langle x, \alpha\rangle + \langle A\beta, \alpha\rangle - H(\alpha)] \\
=&sup_{\alpha\in R^N}[\langle x, \alpha\rangle + \langle \beta, B^K\alpha\rangle - H(\alpha)]
\end{aligned}$$

since for $i \in K$, A^T and B^K have a same vector on i^{th} row and $\beta_j = 0$, for $j \in K^c$.

$$=sup_{\xi^K\in R^N}[\langle -x, \xi^K\rangle + \langle \beta, \xi^K\rangle - H((B^K)^{-1}\xi^K)],$$

since $\xi^K_i = -\alpha_i$ for $i \in K^c$.

$$\begin{aligned}
=&sup_{\xi^K\in R^N}[\langle -x + \beta, \xi^K\rangle - H^K(\xi^K)] \text{ by definition of } H^K \\
=&L^K(-x + \beta). \quad \blacksquare
\end{aligned}$$

Due to Lemma 4.2, we know that $L^K(-x+\beta)$ is a strictly convex and continuously differentiable *w.r.t.* β on $R^{N(+)}_{K^c}$ and increases to ∞ as $||\beta|| \to \infty$. By Lemma A.1, there is a unique $\beta^* \in R^{N(+)}_{K^c}$ such that

$$\inf_{\beta\in R^{N(+)}_{K^c}} L^K(-x + \beta) = L^K(-x + \beta^*).$$

The following theorem shows how the infimum outside the integral can be passed under the integral.

Theorem 4.3. *For a given absolutely continuous function $\phi \in D^N_{(+)}[0,T]$ with $\phi(0) = x \in R^N_{(+)}$,*

$$\inf_{u|\theta(\psi)=\phi,\ \sigma(\psi)=u} \int_0^T L(\phi'(t) + Au'(t))dt$$
$$= \int_0^T \inf_{\beta\in R^{N(+)}_{B(\phi(t))^c}} L(\phi'(t) + A\beta)dt,$$

i.e. $\beta_i > 0$ only if $\phi_i(t) = 0$.

Proof: Let $T_K = \{t \in [0,T] : \phi_i(t) = 0, \forall i \in K, \phi_j(t) > 0, \forall j \in K^c\}$, where $K \subset \mathbb{Z}_N$. By Lemma A.4, $inf_{\beta\in R^{N(+)}_{K^c}} L^K(-\phi'(t) + \beta)$ is measurable on T_K. And for t on which ϕ' is defined and $\phi(t) \in T_K$, by Lemma 4.1, there is a unique $\beta^*(-\phi'(t))$ such that

$$inf_{\beta\in R^{N(+)}_{K^c}} L^K(-\phi'(t) + \beta) = L^K(-\phi'(t) + \beta^*(-\phi'(t)))$$

and, moreover, $\beta^*(-\phi'(t))$ is *a.e.* defined and measurable on T_K.

Define $\pi^*(t) = \int_0^t \sum_{K\in\overline{B(0)}} \beta^*(-\phi'(t))\chi_{T_K}(s)ds$. Since $\sum_{K\in\overline{B(0)}} \beta^*(-\phi'(t))\chi_{T_K}(s)$ is measurable, $\pi^*(t)$ is well-defined and absolutely continuous.

Let $\psi^* = \phi(t) + A\pi^*(t)$. Then $\theta(\psi^*) = \phi$, $\sigma(\psi^*) = \pi^*$. Therefore,

$$\sum_{K\subset\mathbb{Z}_N} \int_{T_K} \inf_{\beta\in R^{N(+)}_{K^c}} L^K(-\phi'(t) + \beta)dt$$
$$= \sum_{K\subset\mathbb{Z}_N} \int_{T_K} L^K(-\phi'(t) + \pi^{*\prime}(t))dt$$
$$= \sum_{K\subset\mathbb{Z}_N} \int_{T_K} L(-\phi'(t) + A\pi^{*\prime}(t))dt$$
$$= \int_0^T L(\psi^{*\prime}(t))dt.$$

If $\theta(\psi) = \phi$, then there exists nondecreasing function u such that $\psi(t) = \phi(t) + Au(t)$. At t where $\phi'(t)$ is defined, $u_i'(t)$ is nonnegative for all $i \in \mathbb{Z}_n$. Therefore,

$L(\psi^{*\prime}) \le L(\psi'(t))$ a.e., $\forall\ \psi \in \{\psi | \theta(\psi) = \phi\}$. It follows that

$$\int_0^T L(\psi^{*\prime}(t))dt \le \int_0^T L(\psi'(t))dt,\ \forall \psi \in \{\psi | \theta(\psi) = \phi\}.$$

On the other hand, since $\theta(\psi^*) = \phi$,

$$\int_0^T L(\psi^{*\prime}(t))dt \ge \inf_{\psi | \theta(\psi)=\phi,\ \sigma(\psi)=u} \int_0^T L(\psi'(t))dt.$$

Therefore,

$$\begin{aligned}
&\inf_{u | \theta(\psi)=\phi,\ \sigma(\psi)=u} \int_0^T L(\phi'(t) + Au'(t))dt \\
&= \int_0^T \sum_{K \subset \mathbf{Z}_N} \inf_{\beta \in R^{N(+)}_{K^c}} L(\phi'(t) + A\beta)\chi_{T_K}(t)dt \\
&= \int_0^T \inf_{\beta \in R^{N(+)}_{B(\phi(t))^c}} L(\phi'(t) + A\beta)dt.
\end{aligned}$$

The desired result is obtained. ∎

Therefore, for an absolutely continuous $\phi \in D^N_{(+)}[0, T]$ with $\phi(0) = x$,

$$J_{[0,T]}(\phi) = \int_0^T \inf_{\beta \in R^{N(+)}_{B(\phi(t))^c}} L(\phi'(t) + A\beta)dt.$$

Our next goal is to find a closed form expression of the infimum inside the integral. Eventually, we will show that the infimum is same as the expression which was found in Dupuis [9]. This is the content of Theorem 4.4 below. We will use the same notations as Dupuis et. al. used in the same paper.

Definition, Notations, and Miscellaneous Notions:.

For a subset S of $\{1, \ldots.., N\}$ i.e. $S \in \overline{\{1, \ldots.., N\}} = \overline{\mathbb{Z}}_N$, choose any point x *s.t.* $\{x: B(x) = S\}$. For $\alpha \in R^N$, define

$$H(S, \alpha) = \sum_{v \in V} \lambda_v(x)(exp(\langle \alpha, v \rangle) - 1).$$

and its Legendre transformation

$$L(S, \beta) = sup_{\alpha \in R^N}[\langle \alpha, \beta \rangle - H(S, \alpha)], \ \beta \in R^N.$$

Obviously, if x in the interior of $R^N_{(+)}$, $H(B(x), \alpha) = H(\alpha)$ which is defined in (3.2).

We sometimes write $H(x, \alpha)$ for $H(B(x), \alpha)$.

Next, let

$$h(x, \alpha) = \vee_{S \subset B(x)} H(S, \alpha),$$

$$l(x, \beta) = sup_{\alpha \in R^N}[\langle \alpha, \beta \rangle - h(x, \alpha)].$$

Define $H^K(S, \xi) = H^K(S, (B^K)^{-1}\xi^K)$.

The main result of this section is in the next theorem.

THEOREM 4.4. *Let $K \subset \mathbb{Z}_N$. Let $x \in R^N_{(+)}$ with $B(x) = K$ and ν be a N dimensional vector with $\nu_i = 0$ for all $i \in K$. Then*

$$\inf_{\beta \in R^{N(+)}_{K^c}} L(\nu + A\beta) = l(x, \nu).$$

Its proof consists of two parts. In the following two lemmas we prove each part.

LEMMA 4.5. *Let $K \subset \mathbb{Z}_N$. For x in the nonnegative orthant of $R^N (x \in R^N_{(+)})$ with $B(x) = K$ and ν be a N dimensional vector with $\nu_i = 0$ for $i \in K$. Then*

$$\inf_{\beta \in R^{N(+)}_{K^c}} L(\nu + A\beta) \leq l(x, \nu)$$

Proof: By Lemma 4.2,

$$\inf_{\beta\in R^{N(+)}_{K^c}} L(\nu + A\beta) = \inf_{\beta\in R^{N(+)}_{K^c}} L^K(-\nu+\beta)$$

We will base our argument on the properties of L^K. Since $L(\nu)$ is a strictly convex and continuously differentiable *w.r.t.* ν on R^N and increases to ∞ as $|\nu| \to \infty$, $L^K(-\nu+\beta)$ is a strictly convex and continuously differentiable *w.r.t.* β on R^N and increases to ∞ as $|\beta| \to \infty$. By Lemma 4.1, there is a unique $\beta^* \in R^{N(+)}_{K^c}$ such that

$$\inf_{\beta\in R^{N(+)}_{K^c}} L^K(-\nu+\beta) = L^K(-\nu+\beta^*).$$

By Lemma A.2, we know that for each $i \in K$,

$$\frac{\partial L^K}{\partial \beta_i}(-\nu+\beta^*) \geq 0 \textit{ and } \frac{\partial L^K}{\partial \beta_i}(-\nu+\beta^*)\cdot\beta_i^* = 0. \tag{4.1}$$

By the property of Legendre transformation(see (1.4)), we know that there is a unique ξ^{K*} such that $L^K(-\nu+\beta^*) = \langle -\nu+\beta^*, \xi^{K*}\rangle - H^K(\xi^{K*})$ and that

$$\frac{\partial L^K}{\partial \beta_i}(-\nu+\beta^*) = \xi_i^{K*}$$

for each $i \in \mathbb{Z}_N$. Therefore, (4.1) can be rewritten as

$$\xi_i^{K*}\cdot\beta_i^* = 0,\ \xi_i^{K*} \geq 0. \tag{4.2}$$

It follows that

$$\begin{aligned} L(\nu + A\beta^*) &= L^K(-\nu+\beta^*) \\ &= \langle -\nu+\beta^*, \xi^{K*}\rangle - H^K(\xi^{K*}) \\ &= \langle -\nu, \xi^{K*}\rangle - H^K(\xi^{K*}) \end{aligned}$$

Suppose $B(\beta^*)\cap K = S_o$. Since $\beta^* \in R^{N(+)}_{K^c}$, by (4.2), we know that

$$\xi_i^{K*}\begin{cases} \geq 0, & i \in S_o \\ = 0, & i \in B(\beta^*)^c, \end{cases} \tag{4.3}$$

since $\beta_i^* > 0$ for $i \in B(\beta^*)^c$.

Now, we partition H^K into two summation terms to see its properties more explicitly. (Recall that $V_K = \{v(i) : i \in K\}$. In particular, $V_{\mathbf{Z}_N} = \{v(i) : i = 1, \cdots, N\}$.)

$$H^K(\xi^{K*}) = \sum_{v\in\ (V/V_K)} \lambda_v(exp(\langle (B^K)^{-1}\xi^{K*}, v\rangle) - 1) + \sum_{v\in\ V_K} \lambda_v(exp(\langle (B^K)^{-1}\xi^{K*}, v\rangle) - 1).$$

For each $i \in K$, $\langle (B^K)^{-1}\xi^{K*}, v(i)\rangle = \langle \alpha^*, v(i)\rangle = \xi_i^{K*}$, where $B^K\alpha^* = \xi^{K*}$. Thus

$$H^K(\xi^{K*}) = \sum_{v\in\ (V/V_K)} \lambda_v(exp(\langle (B^K)^{-1}\xi^{K*}, v\rangle) - 1) + \sum_{i\in\ K} \lambda_{v(i)}(exp(\xi_i^{K*}) - 1).$$

(4.3) implies that for $i \in B(\beta^*)^c$, $exp(\xi_i^{K*}) - 1 = 0$. Therefore,

$$\sum_{i\in K} \lambda_{v(i)}(exp(\xi^*) - 1) = \sum_{i\in\ K/B(\beta^*)^c} \lambda_{v(i)}(exp(\xi_i^{K*}) - 1),$$

and so

$$H^K(\xi^*) = H^K(B(\beta^*)^c, \xi^{K*}).$$

Therefore,

$$L^K(-\nu + \beta^*) = \langle -\nu, \xi^{K*}\rangle - H^K(B(\beta^*)^c, \xi^{K*}). \tag{4.4}$$

For all $i \in K$, $\xi_i^{K*} \geq 0$ and, therefore, $\exp(\xi_i^{K*}) - 1 \geq 0$. It follows that if $S_1 \subset S_2 \subset K$, then $H^K(S_1, \xi^{K*}) \geq H^K(S_2, \xi^{K*})$. Therefore,

$$H^K(\xi^{K*}) = \vee_{S\subset K} H^K(S, \xi^{K*}).$$

It follows from (4.4) that

$$L^K(-\nu + \beta^*) = \langle -\nu, \xi^{K*}\rangle - \vee_{S\subset K} H^K(S, \xi^{K*}).$$

From this result, it follows that

$$L^K(-\nu+\beta^*) \leq sup_{\xi^K}\{\langle -\nu, \xi^K\rangle - \vee_{S\subset K} H^K(S,\xi^K)\}.$$

Restating this in the original variable α, we conclude that

$$\begin{aligned}\inf_{\beta\in R^{N(+)}_{K^c}} L(\nu+A\beta) &\leq sup_\alpha\{\langle \nu,\alpha\rangle - \vee_{S\subset B(x)} H(S,\alpha)\}\\ &= l(x,\nu). \quad \blacksquare\end{aligned}$$

Lemma 4.6. *Let $K \subset \mathbb{Z}_N$. Let $x \in R^N_{(+)}$ with $B(x) = K$ and ν be a N dimensional vector with $\nu_i = 0$ for $i \in K$. Then*

$$\inf_{\beta\in R^{N(+)}_{K^c}} L(\nu+A\beta) \geq l(x,\nu).$$

Proof: By definition,

$$l(x,\nu) = sup_\alpha\{\langle \nu,\alpha\rangle - \vee_{S\subset B(x)} H(S,\alpha)\}. \tag{4.5}$$

Since $\langle \nu,\alpha\rangle - \vee_{S\subset B(x)} H(S,\alpha)$ is continuous *w.r.t.* α and bounded above and decreases to $-\infty$ as $\|\alpha\| \to \infty$, there must exist α^* and $S_o \subset K$ such that

$$sup_\alpha\{\langle \nu,\alpha\rangle - \vee_{S\subset K} H(S,\alpha)\} = \langle \nu,\alpha^*\rangle - H(S_o,\alpha^*). \tag{4.6}$$

Let

$$\Gamma^K = \{\alpha \in R^N | H(S_o,\alpha) = \vee_{S\subset K} H(S,\alpha)\}.$$

Then $\alpha^* \in \Gamma^K$. Hence,

$$\begin{aligned}&\langle \nu,\alpha^*\rangle - H(S_o,\alpha^*)\\ &= sup_{\alpha\in\Gamma^K}\{\langle \nu,\alpha\rangle - H(S_o,\alpha)\}.\end{aligned} \tag{4.7}$$

We know that $\nu_i = 0$ for $i \in K$. This implies that for $i \in K$, changes of α_i do not affect the term $\langle \nu, \alpha \rangle$.

Now, we want to figure out the shape of the set Γ^K. For α in Γ^K,

$$H(S_o, \alpha) \geq H(S, \alpha) \ for \ all \ S \subset K.$$

Therefore, for $i \in S_o$,

$$H(S_o, \alpha) \geq H(S_o/\{i\}, \alpha).$$

So,

$$H(S_o, \alpha) - H(S_o/\{i\}, \alpha) = -\lambda_{v(i)}(exp\langle \alpha, v(i) \rangle - 1) \geq 0.$$

It follows that $\langle \alpha, v(i) \rangle \leq 0$ for each $i \in S_o$.

For $i \in S_o^c \cap K$,

$$H(S_o, \alpha) \geq H(S_o \cup \{i\}, \alpha).$$

So,

$$H(S_o, \alpha) - H(S_o \cup \{i\}, \alpha) = \lambda_{v(i)}(exp\langle \alpha, v(i) \rangle - 1) \geq 0.$$

This implies that $\langle \alpha, v(i) \rangle \geq 0$ on Γ^K for $i \in S_o^c \cap K$. Therefore,

$$\Gamma^K \subset \{\alpha \in R^N : \langle \alpha, v(i) \rangle \leq 0 \ for \ i \in S_o, \ \langle \alpha, v(i) \rangle \geq 0 \ for \ i \in S_o^c \cap K\}.$$

On the other hand, for any $S \subset K$,

$$H(S_o, \alpha) \geq H(S, \alpha)$$

if and only if

$$\sum_{i \in S_o \cap S^c} \lambda_{v(i)}(exp\langle \alpha, v(i) \rangle - 1) \leq \sum_{i \in S_o^c \cap S} \lambda_{v(i)}(exp\langle \alpha_i, v(i) \rangle - 1).$$

However, for any α with $\langle \alpha, v(i) \rangle \leq 0$ for $i \in S_o$ and $\langle \alpha, v(i) \rangle \geq 0$ for $i \in S_o^c \cap K$, the above equality holds for all $S \subset K$. Therefore,

$$\{\alpha \in R^N : \langle \alpha, v(i) \rangle \leq 0 \ for \ i \in S_o, \ \langle \alpha, v(i) \rangle \geq 0 \ for \ i \in S_o^c \cap K\} \subset \Gamma^K.$$

It follows that

$$\Gamma^K = \{\alpha \in R^N : \langle \alpha, v(i)\rangle \leq 0 \ for \ i \in S_o, \ \langle \alpha, v(i)\rangle \geq 0 \ for \ i \in S_o^c \cap K\}.$$

We know that

$$H(S_o, \alpha) = \sum_{v \in V/V_{S_o}} \lambda_v(exp\langle \alpha, v\rangle - 1).$$

By change of variables, we obtain

$$H^K(S_o, \xi^K) = \sum_{v \in V/V_{S_o}} \lambda_v(exp\langle (B^K)^{-1}\xi^K, v\rangle - 1).$$

Now, we want to show that for any $v \notin V_{S_o}$ and $\langle B^{-1}\xi^K, v\rangle$ is a linear combination of ξ_i^K's with negative coefficient for those $i \in S_o$. It may be explained as follows;

Case 1. If $v = v(j)$ for $j \in K/S_o$, $\langle (B^K)^{-1}\xi^K, v\rangle = \xi_j^K$. j is not in S_o.

Case 2. If $v \in V/V_{\mathbb{Z}_N}$, $\langle (B^K)^{-1}\xi^K, v\rangle = \alpha_k$ for some $k \in \mathbb{Z}_N$.

Case 3. If $v \in V_{\mathbb{Z}_N}/V_K$, $\langle (B^K)^{-1}\xi^K, v\rangle = \alpha_k - \xi_j^K$, where some $k \in \mathbb{Z}_N$. $j \notin S_o$.

Therefore, if α_k can be expressed as a linear combination of ξ_i^K's with negative coefficients for $k \in K$, then the explanation will be completed.

So, it suffices to show that for any $k \in \mathbb{Z}_N$, α_k can be expressed as a linear combination of ξ_i^K's with negative coefficients. Due to the topological structure of OCOS system, for a customer to go out of the system starting from the station k , he has to pass a fixed sequence of stations, say $k_1, \cdots, k_n$. If at least one of $k_1, \cdots, k_n$ is in K^c and k_l is the first one among them, then $-\alpha_k = (\alpha_{k_1} - \alpha_k) + (\alpha_{k_2} - \alpha_{k_1}) + \cdots + (\alpha_{k_l} - \alpha_{k_{l-1}}) - \alpha_{k_l} = \xi_k^K + \xi_{k_1}^K + \cdots + \xi_{k_{l-1}}^K + \xi_{k-l}^K$. If none of them are in K^c, $-\alpha_k = \xi_k^K + \xi_{k_1}^K + \cdots + \xi_{k_{n-1}}^K + \xi_{k_n}^K$. Therefore, every α_k can be expressed as a linear combination of ξ_i^K's with negative coefficients. Therefore, for $i \in S_o$, as ξ_i^K increases, $H^K(S_o, \xi^K)$ is nonincreasing.

Our hypothesis that $v_i = 0$ for $i \in K$, and the fact that $\alpha_i = -\xi_i^K$ for such i, implies that

$$
\begin{aligned}
&sup_{\alpha \in \Gamma^K}\{\langle \nu, \alpha\rangle - H(S_o, \alpha)\} \\
=&sup_{\xi^K \in C^K}\{\langle \nu, -\xi^K\rangle - H^K(S_o, \xi^K)\},
\end{aligned}
\tag{4.8}
$$

where

$$
\begin{aligned}
C^K =&\{\xi^K : \xi^K = B^K\alpha,\ \alpha \in \Gamma^K\} \\
=&\{\xi^K \in R^N : \xi_i^K \leq 0 \ for\ i \in S_o,\ \xi_i^K \geq 0\ for\ i \in S_o^c \cap K\}.
\end{aligned}
$$

However, changes of ξ_i^K for $i \in K$ do not affect $\langle \nu, -\xi^K\rangle$ and $H^K(S_o, \xi^K)$ does not increase as ξ_i^K increases for $i \in S_o$. It follows that

$$
\begin{aligned}
&sup_{\xi^K \in C^K}\{\langle -\nu, \xi^K\rangle - H^K(S_o, \xi^K)\} \\
=&sup_{\xi^K \in C^o}\{\langle -\nu, \xi^K\rangle - H^K(S_o, \xi^K)\},
\end{aligned}
\tag{4.9}
$$

where

$$
C^o = \{\xi^K \in R^N : \xi_i^K = 0\ for\ i \in S_o,\ \xi_i^K \geq 0\ for\ i \in S_o^c \cap K\}.
$$

If $\xi_i^K = 0$ for all $i \in S_o$, $H^K(S_o, \xi^K) = H^K(\xi^K)$. Therefore, from (4.5)-(4.9),

$$
l(x, \nu) \leq sup_{\xi^K \in C^o}\{\langle -\nu, \xi^K\rangle - H^K(\xi^K)\}.
$$

When $\beta \in R_{K^c}^{N(+)}$ and $\xi^K \in C^o$, $\langle \beta, \xi^K\rangle = \sum_{i \in K} \beta_i \langle \alpha, v(i)\rangle \geq 0$. Therefore,

$$
l(x, \nu) \leq sup_{\xi^K \in C^o}\{\langle -\nu + \beta, \xi^K\rangle - H^K(\xi^K)\},
$$

and so

$$
l(x, \nu) \leq sup_{\xi^K \in R^N}\{\langle -\nu + \beta, \xi^K\rangle - H^K(\xi^K)\} = L^K(-\nu + \beta),
$$

for all $\beta \in R_{K^c}^{N(+)}$. It follows that

$$
l(x, \nu) \leq \inf_{\beta \in R_{K^c}^{N(+)}} L^K(-\nu + \beta).
$$

By Lemma 4.2,

$$
\begin{aligned}
&L^K(-\nu+\beta)\\
&=sup_{\xi^K\in R^N}[\langle -\nu+\beta,\xi^K\rangle - H^K(\xi^K)]\\
&=sup_{\xi^K\in R^N}[\langle -\nu,-\alpha\rangle + \langle \beta,\xi^K\rangle - H((B^K)^{-1}\xi^K)],
\end{aligned}
$$

since $\nu_i=0$ for $i\in K$ and $\xi^K_j=-\alpha_j$ for $j\in K^c$.

$$=sup_{\alpha\in R^N}[\langle \nu,\alpha\rangle + \langle \beta, B^K\alpha\rangle - H(\alpha)],$$

by definition of H^K.

$$
\begin{aligned}
&=sup_{\alpha\in R^N}[\langle \nu,\alpha\rangle + \langle (B^K)^T\beta,\alpha\rangle - H(\alpha)]\\
&=sup_{\alpha\in R^N}[\langle \nu,\alpha\rangle + \langle A\beta,\alpha\rangle - H(\alpha)],
\end{aligned}
$$

since $\beta_i=0$ for $i\in K^c$ and the i^{th} column of A is equal to that of $(B^K)^T$.

$$=L(\nu+A\beta).$$

It follows that

$$l(x,\nu)\le \inf_{\beta\in R^{N(+)}_{K^c}} L(\nu+A\beta). \quad \blacksquare$$

We now summarize the results of this section. First, the application of the contraction principle to the Skorohod map leads to the fact that the family of probability measures P^ϵ, governing the rescaled queue length processes $\{X^\epsilon(t)\}$, obeys the large deviation principle with rate function $J_{[0,T]}(\cdot,\cdot)$ where

$$
\begin{aligned}
&J_{0,T}(x,\phi)\\
&=\begin{cases}
\int_0^T \inf_{\beta\in R^{N(+)}_{B(\phi(t))^c}} L(\phi'(t)+A\beta)dt, & if\ \phi\ is\ an\ absolutely\ continuous\ function,\\
 & its\ integral\ is\ convergent,\ and\ \phi(0)=x\\
\infty, & o.w..
\end{cases}
\end{aligned}
$$

Second, based on Theorem 4.4, we can put the rate function $J_{[0,T]}(x,\phi)$ into a closed form expression.

THEOREM 4.7. *For an absolutely continuous function $\phi(t) \in R^N_{(+)}$ for all t on $[0,T]$ with $\phi(0) = x \in R^N_{(+)}$,*

$$J_{[0,T]}(x,\phi) = \int_0^T l(\phi(t), \phi'(t))dt,$$

when it is convergent.

PROOF: If $i \in B(\phi(t))$, $\phi_i'(t) = 0$ for almost every t on $[0,T]$, Theorem 4.4 implies the desired result. ■

Finally, the LDP of $\{P \circ (X^\epsilon)^{-1}\}$ is in the following form;

THEOREM 4.7.

(1) For each $c < \infty$, the set $\cup_{x \in O}\{\phi : J_{[0,T]}(x,\phi) \leq c\}$ is a compact set in $D^N[0,T]$ for any compact set $O \subset R^N$, for all $0 < T < \infty$.

(2) For each closed set $F \subset D^N[0,T]$,

$$\limsup_{\epsilon \to 0} \epsilon log P[X^\epsilon(.) \in F | X^\epsilon(0) = x] \leq -\inf_{\phi \in F} J_{[0,T]}(x,\phi).$$

(3) For each open set $G \subset D^N[0,T]$,

$$\liminf_{\epsilon \to 0} \epsilon log P[X^\epsilon(.) \in G | X^\epsilon(0) = x] \geq -\inf_{\phi \in G} J_{[0,T]}(x,\phi).$$

4.2 A Failure of the Skorohod Problem Approach.

In this section, we want to show why the OCOS structure is critical for the Skorohod Problem approach. In fact, in a very simple queueing model with a non-OCOS structure, we can show that there is no continuous Skorohod map which transforms its potential process into the queue length process. The following is the example.

The model of interest consists of two stations. In this model, a customer comes in to the first station and waits to be serviced. Interarrival times to the system from the (system) exterior are independent and exponentially distributed with a positive parameter λ_1. After being serviced, he moves on to the second station. Service times at the first station are independent and exponentially distributed with a positive parameter λ_2. The service times at the second station are independent and exponentially distributed with a positive parameter $\lambda_3 + \lambda_4$. After being serviced at the second station, the customer decides whether to go out of system or to go back to the first station. On average, the probability of leaving the system after being serviced at the second station is $\dfrac{\lambda_3}{\lambda_3 + \lambda_4}$ and that of going back to the first station is $\dfrac{\lambda_4}{\lambda_3 + \lambda_4}$. The small time increment probability of the queue length process is as follows;
when $v^1 = (1,0)^T$, $v^2 = (-1,1)^T$, $v^3 = (0,-1)^T$, and $v^4 = (1,-1)^T$,

$$P_r[X(t+h) - X(t) = v | X(t) = x] = \begin{cases} \lambda_i(x)h + o(h), & v = v^i \\ 1 - \sum_{i=1}^4 \lambda_i(x)h + o(h), & v = (0,0)^T \end{cases},$$

where $\lambda_i(x) = \begin{cases} \lambda_i, & v^i_j \neq -1, \ \forall j \in B(x) \\ 0, & o.w.. \end{cases}$

However, The counterpart of the potential process has the following conditional probability structure.

$$\begin{aligned} P_r[Y(t+h) - Y(t) = v | X(t) = X] &= P_r[Y(t+h) - Y(t) = v] \\ &= \begin{cases} \lambda_i h + o(h), & v = v^i \\ 1 - \sum_{i=1}^4 \lambda_i h + 0(h), & v = (0,0)^T \end{cases} \end{aligned}$$

Suppose there is a Skorohod map θ for this model similar to the one in the case of the OCOS structure (i.e. $\theta(Y)$ has the same distribution as X). Then $\theta(Y)$ should reflect everywhere on the boundary of $R^2_{(+)}$. It implies that, for a given trajectory $Y(w)$ of the potential process Y, if $Y(w,t) - Y(w,t^-) = v$ and $\theta(Y(w))(t^-) = x$ and if $x_i = 0$ and the i^{th} element of v, v_i, is less than 0, then $\theta(Y(w))(t) - \theta(Y(w))(t^-)$ should be equal to 0.

In order to achieve this reflection on the boundary with $x_1 = 0$, we have to have two types of reflections. This makes it impossible to find a continuous Skorohod map for this model. We show it with examples in the following.

We choose two specific trajectories $y_1(\cdot)$, $y_2(\cdot)$ of the potential process and define $y_i^m(t) = \dfrac{1}{n} y_i(nt)$. These two will converge to the same continuous function as n goes to infinity. But $\theta(y_1^\epsilon)(t)$ and $\theta(y_2^\epsilon)(t)$ will converge to two distinct functions. It means θ cannot be a continuous mapping.

$$\text{Let } \; y_1(t) = \begin{cases} 0 & [0,1) \\ v_4 & [1,3) \\ 2v_4 & [3,5) \\ \cdots & \end{cases}$$
$$= \sum_{n=1}^{\infty} \chi_{[2n-1,\ 2n+1)}(t) v_4$$

$$\text{Let } \; y_2(t) = \begin{cases} 0, & [0,1) \\ v_1, & [1,2) \\ v_1 + v_3, & [2,3) \\ 2v_1 + v_3, & [3,4) \\ 2v_1 + 2v_3, & [4,5) \\ \cdots & \end{cases}$$
$$= \sum_{n=1}^{\infty} \chi_{[2n-1,\ 2n)}(t) v_1 + \sum_{n=1}^{\infty} \chi_{[2n,\ 2n+1)}(t) v_3.$$

As n goes to ∞, $y_1^n(t)$ and $y_2^n(t)$ converge to a function $f(t) = \begin{bmatrix} \frac{1}{2}t \\ -\frac{1}{2}t \end{bmatrix}$, uniformly.

For all n, $\theta(y_1^n)(t) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and

$$\theta(y_2^n)(t) = \sum_{n=1}^{\infty} \frac{1}{n} \chi_{[2n-1,\ 2n)}(nt) v_1.$$

As n goes to ∞,

$$\theta(y_2^n)(t) \text{ converges to } g(t) = \begin{bmatrix} \frac{1}{2}t \\ 0 \end{bmatrix}, \text{ uniformly on any bounded interval.}$$

Therefore, θ can not be a continuous mapping.

To push the LDP of the potential process through θ, the continuity of Skorohod map is a critical condition. Due to this reason, our approach has not been successful in cases of multiple reflections.

Appendix

Lemma A.1. *If S is a closed convex subset of R^N, $g(x)$ is strictly convex and continuous on R^N, and $g(x) \to \infty$ as $\|x\| \to \infty$, then there exist a unique point $x^* \in S$ such that $\inf_{x \in S} g(x) = g(x^*)$.*

Proof: Since $g(x) \to \infty$ as $\|x\| \to \infty$, $g(x)$ has the greatest lower bound. Then the continuity of g and the closedness of S guarantee the existence of an optimal point $x^* \in S$ such that $\inf_{x \in S} g(x) = g(x^*)$.

Suppose that there are two optimal points x_1^*, x_2^*. The strict convexity of g and convexity of S implies that there is a point $x' = \frac{x_1^* + x_2^*}{2}$ in S such that $\inf_{x \in S} g(x) > g(x')$. This is a contradiction. Therefore, x^* is unique. ∎

Under the same conditions as Lemma A.1, the Kuhn-Tucker condition (cf. Theorem 7.1.3 in Karlin [17]) and Lemma A.1 implies the following lemma.

Lemma A.2. *Let $R^N_{(+)}$. be the nonnegative orthant of R^N. Provided that $g(x)$ has the same properties as in Lemma A.1 and is continuously differentiable, then $\inf_{x \in R^N_{(+)}} g(x)$ is achieved at a unique $x^* \in R^N_{(+)}$ iff $[\frac{\partial g}{\partial x_i}|_{x^*}] \geq 0$ and $[\frac{\partial g}{\partial x_i}|_{x^*}] \cdot x^* = 0$.*

Lemma A.3. *If $f(x)$ is a real valued and convex function and $\{x : f(x) < \infty\} = R^N$, then $f(x)$ is continuous.*

See B.4 in Karlin [13] for a proof.

Lemma A.4. *Let $f(x)$ be a convex continuous function on R^N and bounded below. Let C be a convex subset of R^N. Define $g(x) = \inf_{c \in C} f(x + c)$, for $x \in R^N$. Then g is convex continuous on R^N.*

PROOF: For any c_1, $c_2 \in C$ and $t \in [0,1]$, $tc_1 + (1-t)c_2 \in C$ and

$$\begin{aligned} g(tx + (1-t)y) &\leq f(tx + (1-t)y + tc_1 + (1-t)c_2) \\ &= f(t(x+c_1) + (1-t)(y+c_2)) \\ &\leq tf(x+c_1) + (1-t)f(y+c_2) \end{aligned}$$

Therefore, $g(tx + (1-t)y) \leq tg(x) + (1-t)g(y)$. It follows that g is convex.

$f(x)$ has a lower bound. Then, obviously, g is bounded below. Moreover, since f is finite on R^N, g is finite on R^N. Therefore, by the Lemma A.3, g is continuous on R^N. ∎

Bibliography

[1] P. Billingsley. *Probability and Measure.* John Wiley, 1979.

[2] C. Costantini. *The Skorohod Oblique Reflection Problem and a Diffusion Approximation for a Class of Transport Processes.* PhD thesis, University of Wisconsin, Madison, 1987.

[3] M. Cottrell, J.-C. Fort, and G. Malgouyres, Large Deviations and Rare Events in the Study of Stochastic Algorithms. *IEEE Trans. on Auto. Control.* 28:907-920, 1983.

[4] H. Cramer. On a New Limit Theorem in the Theory of Probability. Colloquium on the Theory of Probability, Herman, Paris, 1937.

[5] M. G. Crandall and P.-L. lions. Viscosity Solutions of Hamilton-Jacobi Equations. *Trans. Amer. Math. Soc..* 277:1-42, 1983.

[6] P. Dupuis, H. Ishii, and H. M. Soner. A Viscosity Solution Approach to the Asymptotic Analysis of Queueing Systems. *The Annals of Probability.* submitted.

[7] P. Dupuis, R.S. Ellis, and A. Weiss. Large Deviations for Markov Processes with Discontinuous Statistics, I: General Upper Bounds. Technical report. University of Massachusetts, 1989.

[8] P. Dupuis and H. Ishii, On When the Solution to the Skorohod Problem is Lipschitz Continuous, with Applications. *Stochastic and Stochastic Reports.* 35:31-62, 1991.

[9] P. Dupuis. Large Deviation Estimates for Queueing Systems. *Conference Proc..* Imperial College workshop, to appear.

[10] S.N. Ethier and T.G. Kurtz. *Markov Processes: Characterization and Convergence.* John Wiley, 1986.

[11] M. R. Frater and B. D. O. Anderson. Fast Estimation of the Statistics of Excessive Backlogs in Tandem Networks of Queues. *Australian T. R..* 23:49-54, 1989.

[12] M. I. Freidlin and A. D. Wentzell. *Random Perturbation of Dynamical Systems.* Springer, Berlin, New York, 1984.

[13] E. Gelenbe and G. Pujolle. *Introduction to Queueing Networks.* John Wiley, 1987.

[14] Gene H. Golub and Charles F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, 1984

[15] A. G. Greenberg and A. Weiss. An Analysis of Aloha System via Large Deviations. preprint. AT & T Bell Laboratories. July, 1986

[16] J. M. Harrison and M. I. Reiman. Reflected Brownian Motion on an Orthant. *The Annals of Probability.* 9:302-308, 1981.

[17] S. Karlin. *Mathematical Methods and Theory in Games, Programming, and Economics.* Addison-Wesley, 1962.

[18] C. Knessl, B.J. Matkowsky, Z. Schuss, C. Tier. The Two Repairman Problem: A Finite Source M/G/2 Queue. *SIAM J. Applied Math..* 47:367-398, 1987.

[19] C. Knessl, B.J. Matkowsky, Z. Schuss, C. Tier. An Asymptotic Theory of Large Deviations for Markov Jump Processes. *SIAM J. Applied Math..* 46:1006-1028, 1985.

[20] C. Knessl, B.J. Matkowsky, Z. Schuss, C. Tier. On the Performance of State-Dependent Single Server Queues. *SIAM J. Applied Math..* 46:657-698, 1986.

[21] C. Knessl, B.J. Matkowsky, Z. Schuss, C. Tier. Asymptotic Analysis of a State-Dependent M/G/1 Queueing System. *SIAM J. Applied Math..* 46:483-656, 1986.

[22] G. Leitman. *Calculus of Variation and Optimal Control.* Plenum Press, 1983.

[23] S. Parekh and J. Warland. A Quick Simulation Method for Excessive Backlogs in Networks of Queues. *IEEE Trans. on Auto. Control,* 34:54-66, 1989.

[24] R. T. Rockafeller. *Convex Analysis.* Princeton University Press, 1970.

[25] H. Tanaka. Stochastic Differential Equations with Reflecting Boundary Conditions

in Convex Regions. *Hiroshima Math. J.*. 9:163-177, 1979.

[26] S. R. S. Varadhan, Large Deviations and Applications, SIAM, 1984.

[27] A. F. Veinott. Discrete Dynamic Programming with Sensitive Discount Optimality Criteria. *Ann. Math. Statist.*. 40:1635-1660, 1969.

[28] Jean Warland. *Introduction to Queueing Networks.* Prentice-Hall, New Jersey, 1989.

[29] A. Weiss. A New Technique for Analyzing Large Traffic Systems. *Adv. Appl. Prob.*. 18:506-532, 1986.

[30] A. D. Wentzel. Rough Limit Theorems on Large Deviations for Markov Stochastic Processes. I. *Theo. Prob. Appl.*. 21:227-243, 1976.

[31] R. W. Wolff. *Stochastic Modelling and the Theory of Queues.* Prentice-Hall, New Jersey, 1989.

Vita

Young Wook Park was born in September 28, 1956 in Seoul, Korea. He received his B.A. in English literature from Hankuk University of Foreign Studies in Feb. 1980. In 1980, he was admitted to the department of economics, Graduate School of Yonsei University, Seoul, Korea, as a graduate student. While he was in Yonsei University, his major interest was econometrics. After he finished his master's degree in 1982, he became an instructor in the department of statistics, Yonsei University. From Jan. 1983 to Aug. 1984, he worked for Korean government as a econometrician. In 1984, he came to Virginia Polytechnic Institute and State University to study in the department of statistics. After two and a half years of studying statistics, he became interested in probability theory and joined the department of mathematics in 1987. Since then, he has been a graduate student in the department of mathematics. While he was in the department of statistics, he received his master's degree in statistics and passed the Ph.D qualifying examination and completed the course work for Ph.D in statistics.

LARGE DEVIATION THEORY FOR QUEUEING SYSTEMS

Young Wook Park

ABSTRACT

Consider a Markov jump process , X(t), with a nonnegative state space as a model for a queueing system. The motivation of this study is about useful estimates of system performance. For example, in a system with finite queues, the probability of the system of queues going from an empty state to a state in which the population of at least one queue reaches a large number before becoming empty again is one and the typical sample trajectory of this event is another. To answer these questions, we establish the large deviation principle (LDP) for an appropriate class of queueing processes. The model of our concern is the Jackson network which has a tree-type topological structure.

Under carefully designed conditions, the LDP for a time homogeneous Markov process has been well established by Wentzel. However, mainly due to the nonnegativity constraint, the queue length process, X(t), of our model does not satisfy the assumed conditions. As a detour, we define the "potential process", $Y(t)$, which allows the negativity in state space in the way that even if a queue is empty, the server in the empty queue is working with a same rate as if the queue is not empty. Therefore, each $Y_i(t)$ can be expressed as the difference of the accumulated number of customers who came to station i and the accumulated number of services, done in station, i, up to time t. Then the scaled processes, $Y^\epsilon(t) \equiv \epsilon Y(\frac{t}{\epsilon})$, obeys LDP with a

certain rate function, $I_{[0,T]}(x, \phi)$, i.e.

$$P[Y^\epsilon(.) \in B | Y^\epsilon(0) = x] \approx exp[-\frac{1}{\epsilon} \inf_{\phi \in B} I_{[0,T]}(x, \phi)], \ (UTLE)$$

for some $B \subset D^r[0,T] \equiv \{$ right continuous R^r – valued function which has a left limit at every point on $[0,T]\}$. $UTLE$ stands for 'up to logarithmic equivalence'.

By defining an appropriate Skorohod problem, we obtain a continuous mapping θ from D^r to $D^r_{(+)}$ such that $\theta(Y)(t)$ is a version of $X(t)$. Then we "push the LDP of potential process through" θ so that LDP of the queue length process can be achieved. The procedure of 'pushing through' is another principle of the large deviation theory. It is called "contraction principle" [3]. The contraction principle provides the rate function $J_{[0,T]}$ of the LDP for the queue length process and $J_{[0,T]}(\phi) = inf_{\psi|\theta(\psi)=\phi} J_{[0,T]}(\psi)$. That is, when $X^\epsilon \equiv \epsilon X(\frac{t}{\epsilon})$, for an appropriate set $B \subset D^r_{(+)}$,

$$P[X^\epsilon(.) \in B | X^\epsilon(0) = x] \approx exp[-\frac{1}{\epsilon} \inf_{\phi \in B} J_{[0,T]}(x, \phi)]. \ (UTLE)$$

The rate function, $J_{[0,T]}$, is expressed in a closed form.