Super-Exponential Solution for a Retrial Supermarket Model

Quan-Lin Li\textsuperscript{1}  John C.S. Lui\textsuperscript{2}  Yang Wang \textsuperscript{3}
\textsuperscript{1} School of Economics and Management Sciences
Yanshan University, Qinhuangdao 066004, China
\textsuperscript{2} Department of Computer Science & Engineering
The Chinese University of Hong Kong, Shatin, N.T, Hong Kong
\textsuperscript{3} Institute of Network Computing & Information Systems
Peking University, China

January 22, 2013

Abstract

In this paper, we provide a new and effective approach for studying super-exponential solution of a retrial supermarket model with Poisson arrivals, exponential service times and exponential retrial times and with two different probing-server numbers. We describe the retrial supermarket model as a system of differential equations by means of density-dependent jump Markov processes, and obtain an iterative algorithm for computing the fixed point of the system of differential equations. Based on the fixed point, we analyze the expected sojourn time that a tagged arriving customer spends in this system, and use numerical examples to indicate different influence of the two probing-server numbers on system performance including the fixed point and the expected sojourn time. Furthermore, we analyze exponential convergence of the current location of the retrial supermarket model to the fixed point, and apply the Kurtz Theorem to study density-dependent jump Markov process given in the retrial supermarket model, which leads to a Lipschitz condition under which the fraction measure of the retrial supermarket model weakly converges to the system of differential equations. This paper arrives at a new understanding of how the workload probing can help in load balancing jobs in retrial supermarket models.

Keywords: Randomized load balancing, retrial queue; supermarket model, super-
exponential solution, density-dependent jump Markov process, fixed point, sojourn
time, exponential convergence, Lipschitz condition.

1 Introduction

This paper analyzes a retrial supermarket model with Poisson arrivals, exponential service
times and exponential retrial times and with two different probing-server numbers. It
proposes a new analytic approach for discussing retrial queueing networks. This new
approach is based on both differential equation approximation for Markov chains and
limiting structure for population processes, e.g., see Kurtz [16], Wormald [43] and Darling
and Norris [6]. A key result of this paper is an iterative algorithm for computing the
fixed point, which is always important in the study of supermarket models. Based on the
fixed point, this paper uses numerical examples to indicate different influence of the two
probing-server numbers on system performance including the fixed point and the expected
sojourn time.

Recently, a number of companies (e.g., Amazon, Google etc.) are offering cloud computing service and cloud manufacturing technology. This motivates us to study efficient
load balancing in large-scale networks with many computational and manufacturing resources. It is well-known that optimal task assignment is a key design issue when one can
use different resources of large-scale networks. For such task assignment, there are two
main approaches: Push strategies and pull strategies. Push strategies are mainly used in
centralized systems, such as data centers allocate jobs to different processors; while pull
strategies are applied to send work from an overloaded processor to underloaded proces-
sors, or an underloaded resource will ask for work from overloaded resources. In the study
of push strategies, randomized load balancing is an important method, and specifically,
retrial supermarket models provide a crucial task assignment mechanism in randomized
load balancing schemes.

Randomized load balancing, where a job is assigned to a server from a small sub-
set of randomly chosen servers, is very simple to implement. It has surprisingly good
performance (for example, reducing collisions, waiting times, backlogs) in a number of
applications, such as, data centers, hash tables, distributed memory machines, path selec-
tion in networks, and task assignment at web servers. Supermarket models are extensively
used to study randomized load balancing schemes. In the study of supermarket models,
a key result by Vvedenskaya, Dobrushin and Karpelevich [39] indicated that when each Poisson arriving job is assigned to the shortest one of \( d \geq 2 \) randomly chosen queues with exponential service times, the equilibrium queue length can decay doubly exponentially in the limit as the population size \( n \to \infty \). Also, the stationary fraction of queues with at least \( k \) customers is \( \rho^{\frac{k}{d}} \). This is a substantially performance improvement over the case for \( d = 1 \), where the tail of stationary queue length distribution in the corresponding M/M/1 queue is \( \rho^k \). Further, such an exponential improvement is illustrated by Li and Lui [19] for studying doubly exponential solution of the supermarket model with Markovian arrival processes and phase-type service times, and by Luczak and McDiarmid [22] for analyzing the maximum queue length in the supermarket model with Poisson arrivals and exponential service times.

In the past ten years, supermarket models have been formulated by queueing theory as well as Markov processes. Much recurrent research dealt with the supermarket model with Poisson arrival processes and exponential service times by means of density-dependent jump Markov processes, which are applied to describe limiting behavior of the supermarket model as an infinite-size system of differential equations whose fixed point is doubly exponential. Readers may refer to Vvedenskaya, Dobrushin and Karpelevich [39] and Mitzenmacher [28]. Certain generalization of supermarket models has been explored in, for example, studying simple variations by Vvedenskaya and Suhov [40], Mitzenmacher and Vöcking [35], Mitzenmacher [29, 30, 33], Azar, Broder, Karlin and Upfal [3], Vöcking [38] and Mitzenmacher, Richa, and Sitaraman [34]; considering non-Poisson arrivals or/and non-exponential service times by Li, Lui and Wang [20, 21], Li and Lui [19], Bramson, Lu and Prabarker [1] and Li [18]; discussing load information by Mirchandaney, Towsley, and Stankovic [36], Dahlin [3] and Mitzenmacher [32, 34]; mathematical analysis by Graham [11, 12, 13], Luczak and Norris [24] and Luczak and McDiarmid [22, 23]; using fast Jackson networks by Martin and Suhov [26], Martin [25] and Suhov and Vvedenskaya [37].

Retrial queues are an important mathematical model for studying telephone switch systems, digital cellular mobile networks, computer networks and so on. During the last two decades, considerable attention has been paid to the study of retrial queues, for example, by survey papers of Yang and Templeton [44], Falin [8], Kulkarni and Liang [15], Artalejo [1] and Gómez-Corral [10], and by three books of Falin and Templeton [9], Artalejo and Gómez-Corral [2] and Li [17]. Note that the study of retrial queueing networks is difficult and challenging with few available results up to now, this paper proposes a novel
approach for studying a class of important retrial queueing networks.

The main contributions of the paper are threefold. The first one is to provide a novel approach to study a retrial supermarket model with Poisson arrivals, exponential service times and exponential retrial times and with two different probing-server numbers. We describe the retrial supermarket model as a system of differential equations by means of density-dependent jump Markov processes, and give an iterative algorithm for computing the fixed point which is shown to be super-exponential. The second one is to give an iterative algorithm for computing the expected sojourn time by means of the fixed point, and use numerical examples to illustrate different influence of the two probing-server numbers on system performance including the expected sojourn time. The third one is in analyzing exponential convergence of the current location of the retrial supermarket model to the fixed point. Not only does the exponential convergence indicate existence of the fixed point, but it also explains that such a convergent process is very fast. Furthermore, to study limiting behavior of the retrial supermarket model as its population size goes to infinity, we apply the Kurtz Theorem to study density-dependent jump Markov process corresponding to the retrial supermarket model. This leads to a Lipschitz condition under which the fraction measure of the retrial supermarket model weakly converges to the system of differential equations. Different from previous works in the literature of retrial queues, this paper gives a more general analytical approach in the study of retrial queueing networks. We believe our approach can open a new door to performance evaluation of more complicated retrial queueing networks.

The remainder of this paper is organized as follows. In Section 2, we describe a retrial supermarket model with Poisson arrivals, exponential service times and exponential retrial times and with two different probing-server numbers, and derive a system of differential equations based on density-dependent jump Markov processes. In Section 3, we provide an iterative algorithm for computing the fixed point of the system of differential equations, and show that the fixed point is super-exponential. In Section 4, we provide an effective algorithm for computing the expected sojourn time, and use numerical examples to indicate different influence of the two probing-server numbers on system performance including the expected sojourn time. In Section 5, we study exponential convergence of the current location of the retrial supermarket model to the fixed point. In Section 6, we apply the Kurtz Theorem to study density-dependent jump Markov process corresponding to the retrial supermarket model, which leads to a Lipschitz condition under which
the fraction measure of the retrial supermarket model weakly converges to the system of differential equations. Some concluding remarks are given in Section 7.

2 Retrial Supermarket Model Description

In this section, we describe a retrial supermarket model with Poisson arrivals, exponential service times and exponential retrial times and with two different probing-server numbers, and derive a system of differential equations based on density-dependent jump Markov processes.

Let us describe the retrial supermarket model. Customers arrive at a queueing system of \( n > 1 \) servers as a Poisson process with arrival rate \( n\lambda \) for \( \lambda > 0 \), and the service times at each server are exponential with service rate \( \mu > 0 \). There is no waiting space before each server, but there is a total orbit of infinite size for all waiting customers. Each primary arriving customer chooses \( d_1 \geq 1 \) servers independently and uniformly at random from the \( n \) servers. If there is no customer in one of the \( d_1 \) chosen servers, then it enters the idle server immediately and receives service; otherwise it enters the orbit and makes a retrial at a later time. Each retrial arriving customer chooses \( d_2 \geq 1 \) servers independently and uniformly at random from the \( n \) servers, if there is no customer in one of the \( d_2 \) chosen servers, then it enters the idle server immediately and receives service; otherwise it enters the orbit again and makes a retrial at a later time. Returning customers behave independently of each other and are persistent in the sense that they keep making retrials until they receive their requested service. Successive inter-retrial times of each customer in the orbit are exponential with retrial rate \( n\theta \) for \( \theta > 0 \). We assume that all the random variables defined above are independent of each other, and that this queueing system is operating in the region \( \rho = \lambda/\mu < 1 \).

In the retrial supermarket model, Figure 1 provides physical structure of the retrial supermarket model, and Figure 2 shows that each primary arriving customer and each retrial arriving customer can choose \( d_1 \) and \( d_2 \) servers independently and uniformly at random from the \( n \) servers, respectively. We call \( d_1 \) and \( d_2 \) the two probing-server numbers. Note that the two probing-server numbers have different influence on system performance including the fixed point and the expected sojourn time.

The following lemma, which is stated without proof, intuitively provides a sufficient condition under which the retrial supermarket model is stable. Note that this proof can be
Figure 1: Physical structure of the retrial supermarket model

Figure 2: The primary and retrial customers probe the loading of $d_1, d_2 \geq 1$ servers, respectively
given by a simple comparison argument with the queueing system in which each customer queues at a random server (i.e., where \( d = 1 \)). When \( d = 1 \), each server acts like an M/M/1 retrial queue which is stable if \( \rho = \lambda/\mu < 1 \), see Falin and Templeton \[9\] or Artalejo and Gómez-Corral \[2\]. The comparison argument is similar to those in Winston \[42\] and Weber \[41\], thus we obtain two useful results: (1) the shortest queue is optimal due to the assumptions on Poisson arrivals, exponential service times and exponential retrial times; and (2) the size of the longest queue in the retrial supermarket model is stochastically dominated by the size of the longest queue in a set of \( n \) independent M/M/1 retrial queues.

Lemma 1 For the retrial supermarket model with Poisson arrivals, exponential service times and exponential retrial times and with two different probing-server numbers, it is stable if \( \rho = \lambda/\mu < 1 \).

For \( k \geq 0 \), we denote by \( n_k^{(W)}(t) \) the numbers of busy servers with at least \( k \) customers in the orbit at time \( t \geq 0 \), and \( n_k^{(I)}(t) \) the numbers of idle servers with at least \( k \) customers in the orbit at time \( t \geq 0 \). Clearly, \( n_0^{(W)}(t) + n_0^{(I)}(t) = n \) and \( 0 \leq n_k^{(W)}(t), n_k^{(I)}(t) \leq n \) for \( k \geq 0 \).

For \( k \geq 0 \), we write

\[
x_n^{(W)}(k, t) = \frac{n_k^{(W)}(t)}{n},
\]

and

\[
x_n^{(I)}(k, t) = \frac{n_k^{(I)}(t)}{n},
\]

which are the fractions of busy servers and of idle servers with at least \( k \) customers in the orbit at time \( t \geq 0 \). Let

\[
X_n(0, t) = x_n^{(I)}(0, t)
\]

and for \( k \geq 1 \)

\[
X_n(k, t) = \left( x_n^{(W)}(k - 1, t), x_n^{(I)}(k, t) \right),
\]

\[
X_n(t) = (X_n(0, t), X_n(1, t), X_n(2, t), \ldots).
\]

The state of the retrial supermarket model can be described as the vector \( X_n(t) \) for \( t \geq 0 \). Since the arrival process to this queueing system is Poisson and the service or/and retrial times of each customer are all exponential, \( \{X_n(t), t \geq 0\} \) is a Markov process whose state
space is given by
\[ \Omega_n = \{(g_n^{(0)}, g_n^{(1)}, g_n^{(2)}, \ldots) : g_n^{(0)} \in (0, 1), g_n^{(k)} \geq g_n^{(k+1)} \geq 0, \text{ and } ng_n^{(k)} \text{ is a two-dimensional row vector of nonnegative integers for } k \geq 1\} \]

For \( k \geq 0 \), we write
\[ s_k^{(W)} (n, t) = E \left[ x_n^{(W)} (k, t) \right] \]
and
\[ s_k^{(I)} (n, t) = E \left[ x_n^{(I)} (k, t) \right] . \]

Let
\[ S_0 (n, t) = s_0^{(I)} (n, t) \]
and for \( k \geq 1 \)
\[ S_k (n, t) = \left( s_{k-1}^{(W)} (n, t), s_k^{(I)} (n, t) \right) , \]
\[ S (n, t) = (S_0 (n, t), S_1 (n, t), S_2 (n, t) , \ldots) . \]

As shown in Martin and Suhov [26] and Luczak and McDiarmid [22], the Markov process \( \{X_n (t), t \geq 0\} \) is asymptotically deterministic as the population size \( n \to \infty \). Thus, \( \lim_{n \to \infty} E \left[ x_n^{(W)} (k, t) \right] \) and \( \lim_{n \to \infty} E \left[ x_n^{(I)} (k, t) \right] \) always exist by means of the law of large numbers for \( k \geq 0 \). Based on this, for \( k \geq 0 \) we write
\[ S_k^{(W)} (t) = \lim_{n \to \infty} s_k^{(W)} (n, t) , \]
\[ S_k^{(I)} (t) = \lim_{n \to \infty} s_k^{(I)} (n, t) , \]
\[ S_k (t) = \lim_{n \to \infty} S_k (n, t) \]
and
\[ S (t) = (S_0 (t), S_1 (t), S_2 (t) , \ldots) . \]

Let \( X (t) = \lim_{n \to \infty} X_n (t) \). Then it is easy to see from the Poisson arrivals, the exponential service times and the exponential retrial times that \( \{X (t), t \geq 0\} \) is also a Markov process whose state space is given by
\[ \Omega = \left\{ (g^{(0)}, g^{(1)}, g^{(2)}, \ldots) : g^{(0)} \in (0, 1), g^{(k)} \geq g^{(k+1)} \geq 0 \text{ for } k \geq 1 \right\} . \]

If the initial distribution of the Markov process \( \{X_n (t), t \geq 0\} \) approaches the Dirac delta-measure concentrated at a point \( g \in \Omega \), then \( X (t) = \lim_{n \to \infty} X_n (t) \) is concentrated on the
trajectory $S_g = \{ S(t) : t \geq 0 \}$. This indicates a law of large numbers for the time evolution of the fractions of busy servers and of idle servers. Furthermore, the Markov process $\{X_n(t), t \geq 0\}$ converges weakly to the fraction vector $S(t) = (S_0(t), S_1(t), S_2(t), \ldots)$ as the population size $n \to \infty$, or for a sufficiently small $\varepsilon > 0$,

$$\lim_{n \to \infty} P \{|X_n(t) - S(t)| \geq \varepsilon\} = 0,$$

where $\|a\|$ is the $L_\infty$-norm of vector $a$.

The following proposition shows that the three sequences $\{S^{(W)}_k(t) + S^{(I)}_k(t)\}$, $\{S^{(W)}_k(t)\}$ and $\{S^{(I)}_k(t)\}$ are monotone decreasing, while its proof is easy by means of the definition of the fraction measure $S(t)$. Note that the crucial number $K_{d_1,d_2}$ will be indicated in Tables 1 to 3 later.

**Proposition 1** For $k \geq 1$

$$S^{(W)}_{k+1}(t) + S^{(I)}_{k+1}(t) < S^{(W)}_k(t) + S^{(I)}_k(t) < S^{(W)}_0(t) + S^{(I)}_0(t) = 1.$$

At the same time, there exists an integer $K_{d_1,d_2} \geq 0$ such that for $k \geq K_{d_1,d_2}$

$$S^{(W)}_{k+1}(t) < S^{(W)}_k(t)$$

and

$$S^{(I)}_{k+1}(t) < S^{(I)}_k(t).$$

Now, we set up a system of differential equations in order to determine the fraction measure $S(t)$ by means of density-dependent jump Markov processes. To that end, we use Figure 3 to provide a concrete example for $k \geq 1$ to indicate how to derive the system of differential equations.

![Figure 3: State transitions of a retrial queue in the retrial supermarket model](image_url)
Consider the retrial supermarket model with \( n \) servers, and determine the expected change in the number of busy servers over a small time period of length \( dt \). During the time interval \([0, dt]\), the probability that any primary arriving customer joins the orbit with \( k - 1 \) customers is given by \( n\lambda \left[ s_{k-1}^{(W)}(n, t) \right] d^1 t \), the probability that any primary arriving customer enters an idle server with \( k \) customers in the orbit is given by \( n\lambda \left\{ -\left[ s_k^{(W)}(n, t) \right] d^1 + \left[ s_k^{(I)}(n, t) \right] d^1 \right\} dt \), the probability that any retrial arriving customer enters an idle server with \( k + 1 \) customers in the orbit is given by \((k + 1)n\theta \left[ s_{k+1}^{(I)}(n, t) \right] d^2 dt \), and the probability that any customer finishes its required service and leaves this system with \( k \) customers in the orbit is given by \( n \left[ \mu s_k^{(W)}(n, t) \right] dt \).

Therefore, we obtain

\[
\frac{dE[n_k^{(W)}(t)]}{dt} = n\lambda \left[ s_{k-1}^{(W)}(n, t) \right] d^1 - n\lambda \left[ s_k^{(W)}(n, t) \right] d^1 + (k + 1) n\theta \left[ s_{k+1}^{(I)}(n, t) \right] d^2 + n \left[ -\mu s_k^{(W)}(n, t) \right],
\]

which leads to

\[
\frac{ds_k^{(W)}(n, t)}{dt} = \lambda \left[ s_{k-1}^{(W)}(n, t) \right] d^1 - \lambda \left[ s_k^{(W)}(n, t) \right] d^1 + (k + 1) \theta \left[ s_{k+1}^{(I)}(n, t) \right] d^2 + \lambda \left[ s_k^{(I)}(n, t) \right] d^1 - \mu s_k^{(W)}(n, t).
\]

Noting that \( \lim_{n\to\infty}s_k^{(W)}(n, t) \) exists for \( k \geq 0 \) and taking \( n \to \infty \) in both sides of Equation (1), we have

\[
\frac{dS_k^{(W)}(t)}{dt} = \lambda \left[ S_{k-1}^{(W)}(t) \right] d^1 - \lambda \left[ S_k^{(W)}(t) \right] d^1 + (k + 1) \theta \left[ S_{k+1}^{(I)}(t) \right] d^2 + \lambda \left[ S_k^{(I)}(t) \right] d^1 - \mu S_k^{(W)}(t).
\]

Using a similar analysis to Equation (1), we obtain the system of differential equations satisfied by the fraction vector \( S(n, t) = (S_0(n, t), S_1(n, t), S_2(n, t), \ldots) \) as follows:

\[
\frac{ds_0^{(I)}(n, t)}{dt} = -\lambda \left[ s_0^{(I)}(n, t) \right] d^1 + \mu s_0^{(W)}(n, t),
\]

\[
\frac{ds_0^{(W)}(n, t)}{dt} = \lambda \left[ s_0^{(I)}(n, t) \right] d^1 - \lambda \left[ s_0^{(W)}(n, t) \right] d^1 + \theta \left[ s_1^{(I)}(n, t) \right] d^2 - \mu s_0^{(W)}(n, t),
\]

for \( k \geq 1 \)

\[
\frac{ds_k^{(W)}(n, t)}{dt} = \lambda \left[ s_{k-1}^{(W)}(n, t) \right] d^1 - \lambda \left[ s_k^{(W)}(n, t) \right] d^1 + (k + 1) \theta \left[ s_{k+1}^{(I)}(n, t) \right] d^2 + \lambda \left[ s_k^{(I)}(n, t) \right] d^1 - \mu s_k^{(W)}(n, t).
\]
and
\[
\frac{d s_k^{(I)}(n, t)}{dt} = -\lambda \left[ s_k^{(I)}(n, t) \right]^{d_1} - k\theta \left[ s_k^{(I)}(n, t) \right]^{d_2} + \mu s_k^{(W)}(n, t).
\] (6)

Note that \( s_0^{(I)}(n, t) + s_0^{(W)}(n, t) = 1 \).

Since \( \lim_{n \to \infty} S_k(n, t) \) exists for \( k \geq 0 \), taking \( n \to \infty \) in both sides of the system of differential vector equations (3) to (6), we obtain a system of differential vector equations for the fraction vector \( S(t) = (S_0(t), S_1(t), S_2(t), \ldots) \) as follows:

\[
\frac{d S_0^{(I)}(t)}{dt} = -\lambda \left[ S_0^{(I)}(t) \right]^{d_1} + \mu S_0^{(W)}(t),
\] (7)

\[
\frac{d S_0^{(W)}(t)}{dt} = \lambda \left[ S_0^{(I)}(t) \right]^{d_1} - \lambda \left[ S_0^{(W)}(t) \right]^{d_1} + \theta \left[ S_1^{(I)}(t) \right]^{d_2} - \mu S_0^{(W)}(t),
\] (8)

for \( k \geq 1 \)

\[
\frac{d S_k^{(W)}(t)}{dt} = \lambda \left[ S_{k-1}^{(W)}(t) \right]^{d_1} - \lambda \left[ S_k^{(W)}(t) \right]^{d_1} + (k + 1) \theta \left[ S_{k+1}^{(I)}(t) \right]^{d_2} + \lambda \left[ S_k^{(I)}(t) \right]^{d_1} - \mu S_k^{(W)}(t)
\] (9)

and

\[
\frac{d S_k^{(I)}(t)}{dt} = -\lambda \left[ S_k^{(I)}(t) \right]^{d_1} - k\theta \left[ S_k^{(I)}(t) \right]^{d_2} + \mu S_k^{(W)}(t).
\] (10)

Note that \( S_0^{(I)}(t) + S_0^{(W)}(t) = 1 \).

### 3 Super-Exponential Solution

In this section, we provide an iterative algorithm for computing the fixed point of the system of differential equations (7) to (10), and show that the fixed point is super-exponential.

Let \( \pi_0 = \pi_0^{(I)} \) and \( \pi_k = \left( \pi_k^{(W)}, \pi_k^{(I)} \right) \) for \( k \geq 1 \). A row vector \( \pi = (\pi_0, \pi_1, \pi_2, \ldots) \) is called a fixed point of the fraction vector \( S(t) \) if \( \lim_{t \to +\infty} S(t) = \pi \). It is obvious that if \( \pi \) is a fixed point of the fraction vector \( S(t) \), then

\[
\lim_{t \to +\infty} \left[ \frac{d}{dt} S(t) \right] = 0.
\]

Taking \( t \to +\infty \) in both sides of the system of differential equations (7) to (10), together with \( S_0^{(I)}(t) + S_0^{(W)}(t) = 1 \), we obtain a system of nonlinear equations as follows:

\[
\pi_0^{(I)} + \pi_0^{(W)} = 1,
\] (11)
\[-\lambda \left[ \pi_0^{(I)} \right]^{d_1} + \mu \pi_0^{(W)} = 0, \quad (12)\]

\[\lambda \left[ \pi_0^{(I)} \right]^{d_1} - \lambda \left[ \pi_0^{(W)} \right]^{d_1} + \theta \left[ \pi_1^{(I)} \right]^{d_2} - \mu \pi_0^{(W)} = 0, \quad (13)\]

for \(k \geq 1\)

\[\lambda \left[ \pi_k^{(W)} \right]^{d_1} - \lambda \left[ \pi_k^{(W)} \right]^{d_1} + (k + 1) \theta \left[ \pi_{k+1}^{(I)} \right]^{d_2} + \lambda \left[ \pi_k^{(I)} \right]^{d_1} - \mu \pi_k^{(W)} = 0 \quad (14)\]

and

\[-\lambda \left[ \pi_k^{(I)} \right]^{d_1} - k \theta \left[ \pi_k^{(I)} \right]^{d_2} + \mu \pi_k^{(W)} = 0. \quad (15)\]

The following proposition shows that the three sequences \(\left\{ \pi_k^{(W)} + \pi_k^{(I)} \right\}, \left\{ \pi_k^{(W)} \right\}\) and \(\left\{ \pi_k^{(I)} \right\}\) are monotone decreasing, while the proof is easy in terms of Proposition 1. Note that the crucial number \(K_{d_1, d_2}\) will be indicated in Tables 1 to 3 later.

**Proposition 2** For \(k \geq 1\)

\[\pi_{k+1}^{(W)} + \pi_{k+1}^{(I)} < \pi_k^{(W)} + \pi_k^{(I)} < \pi_0^{(W)} + \pi_0^{(I)} = 1.\]

At the same time, there exists an integer \(K_{d_1, d_2} \geq 0\) such that for \(k \geq K_{d_1, d_2}\)

\[\pi_{k+1}^{(W)} < \pi_k^{(W)}\]

and

\[\pi_{k+1}^{(I)} < \pi_k^{(I)}.\]

By using the system of nonlinear equations (11) to (15), we begin to derive the closed-form solution for the fixed point \(\pi = (\pi_0, \pi_1, \pi_2, ...)\). To that end, we denote by \(\eta^* = 1 - \eta_{d_1}\) a positive solution to the nonlinear equation

\[\lambda x^{d_1} + \mu x - \mu = 0 \quad (16)\]

or

\[x = 1 - \rho x^{d_1}.\]

**Proposition 3** There exists the unique positive solution \(\eta^*\) to the nonlinear equation \(\lambda x^{d_1} + \mu x - \mu = 0\) for \(x \in (0, +\infty)\), and \(\eta^* \in (0, 1)\).
**Proof:** If \( \eta^* \) is an any positive solution to the equation \( x = 1 - \rho x^{d_1} \) for \( x \in (0, +\infty) \), then it is clear that \( \eta^* \in (0, 1) \) due to the fact that \( x = 1 - \rho x^{d_1} < 1 \). Let \( G(x) = \rho x^{d_1} + x \) and \( H(x) = 1 \). Since \( G(0) = 0 \) and \( G(1) = 1 + \rho > 1 \), there the two continuous functions: \( y = G(x) \) and \( y = 1 \), must have an intersection point \( \eta^* \in (0, 1) \) by means of the intermediate value theorem, that is, the equation \( x = 1 - \rho x^{d_1} \) must exist a positive solution in the interval \( (0, 1) \). Note that \( \frac{d}{dx} G(x) = \rho d_1 x^{d_1-1} + 1 > 0 \) for \( x \in (0, 1) \), thus \( G(x) \) is strictly monotone increasing for \( x \in (0, 1) \). This shows that the positive solution \( \eta^* \) is unique for the nonlinear equation \( G(x) = H(x) \), or \( x = 1 - \rho x^{d_1} \). This completes the proof.

**Remark 1**

The numerical implementations are used to illustrate that the unique positive solution \( \eta^* \) to the nonlinear equation \( x = 1 - \rho x^{d_1} \) for \( x \in (0, 1) \) has two monotone properties: \( \eta^* \) is a monotone increasing function of \( d_1 \geq 1 \) for a given \( \rho \in (0, 1) \), while it is a monotone decreasing function of \( \rho \in (0, 1) \) for a given \( d_1 \geq 1 \).

When \( d_1 = 3 \), the positive solution is

\[
\begin{align*}
\rho &= 0.1 \quad 0.2 \quad 0.3 \quad 0.4 \quad 0.5 \quad 0.6 \quad 0.7 \quad 0.8 \quad 0.9 \\
\eta^* &= 0.92169 \quad 0.86883 \quad 0.82905 \quad 0.79728 \quad 0.77091 \quad 0.74844 \quad 0.72890 \quad 0.71165 \quad 0.69624
\end{align*}
\]

When \( d_1 = 5 \), the positive solution is

\[
\begin{align*}
\rho &= 0.1 \quad 0.2 \quad 0.3 \quad 0.4 \quad 0.5 \quad 0.6 \quad 0.7 \quad 0.8 \quad 0.9 \\
\eta^* &= 0.93031 \quad 0.88896 \quad 0.85937 \quad 0.83633 \quad 0.81747 \quad 0.80151 \quad 0.78770 \quad 0.77554 \quad 0.76468
\end{align*}
\]

When \( d_1 = 10 \), the positive solution is

\[
\begin{align*}
\rho &= 0.1 \quad 0.2 \quad 0.3 \quad 0.4 \quad 0.5 \quad 0.6 \quad 0.7 \quad 0.8 \quad 0.9 \\
\eta^* &= 0.94387 \quad 0.91643 \quad 0.89785 \quad 0.88375 \quad 0.87236 \quad 0.86281 \quad 0.85457 \quad 0.84734 \quad 0.84089
\end{align*}
\]

When \( d_1 = 20 \), the positive solution is

\[
\begin{align*}
\rho &= 0.1 \quad 0.2 \quad 0.3 \quad 0.4 \quad 0.5 \quad 0.6 \quad 0.7 \quad 0.8 \quad 0.9 \\
\eta^* &= 0.95779 \quad 0.94088 \quad 0.92989 \quad 0.92169 \quad 0.91513 \quad 0.90967 \quad 0.90497 \quad 0.90086 \quad 0.89719
\end{align*}
\]

When \( d_1 = 50 \), the positive solution is

\[
\begin{align*}
\rho &= 0.1 \quad 0.2 \quad 0.3 \quad 0.4 \quad 0.5 \quad 0.6 \quad 0.7 \quad 0.8 \quad 0.9 \\
\eta^* &= 0.97366 \quad 0.96547 \quad 0.96033 \quad 0.95656 \quad 0.95357 \quad 0.95109 \quad 0.94897 \quad 0.94711 \quad 0.94547
\end{align*}
\]

Let

\[
\delta_0 = \rho (1 - \eta_{d_1})^{d_1}
\]
and for \( k \geq 1 \)
\[
\delta_k = \rho \delta_{k-1}^d + \rho \left( \frac{\lambda}{k \theta} \delta_{k-1}^d \right)^{\frac{d_2}{d_1}}
\]

The following theorem provides an iterative algorithm for computing the fixed point which is shown to be super-exponential.

**Theorem 1** The fixed point \( \pi = (\pi_0, \pi_1, \pi_2, \ldots) \) is given by

\[
\begin{align*}
\pi_0^{(I)} &= 1 - \eta d_1, \\
\pi_0^{(W)} &= \delta_0,
\end{align*}
\]

for \( k \geq 1 \)
\[
\begin{align*}
\pi_k^{(I)} &= \left( \frac{\lambda}{k \theta} \delta_{k-1}^d \right)^{\frac{1}{d_2}}, \\
\pi_k^{(W)} &= \delta_k.
\end{align*}
\]

**Proof:** By induction, we can derive the above desired result.

It follows from Equations (11), (12) and (16) that
\[
\lambda \left[ \pi_0^{(I)} \right]^{d_1} + \mu \pi_0^{(I)} - \mu = 0.
\]
This gives
\[
\pi_0^{(I)} = \eta^* = 1 - \eta d_1,
\]
and
\[
\pi_0^{(W)} = \rho (1 - \eta d_1)^{d_1} = \delta_0.
\]

It follows from (13) that
\[
\lambda (1 - \eta d_1)^{d_1} - \lambda \rho d_1 (1 - \eta d_1)^{d_2} + \theta \left[ \pi_1^{(I)} \right]^{d_2} - \mu \rho (1 - \eta d_1)^{d_1} = 0
\]
which follows
\[
\pi_1^{(I)} = \left[ \frac{\lambda}{\theta} \rho d_1 (1 - \eta d_1)^{d_2} \right]^{\frac{1}{d_2}} = \left( \frac{\lambda}{\theta} \delta_{d_1}^0 \right)^{\frac{d_2}{d_1}}.
\]
Together with (15) for \( k = 1 \), this gives
\[
\begin{align*}
\pi_1^{(W)} &= \rho \rho d_1 (1 - \eta d_1)^{d_2} + \rho \left[ \frac{\lambda}{\theta} \rho d_1 (1 - \eta d_1)^{d_2} \right]^{\frac{d_2}{d_1}} \\
&= \rho \delta_{d_0}^d + \left( \frac{\lambda}{\theta} \delta_{d_0}^d \right)^{\frac{d_2}{d_1}} \\
&= \delta_1.
\end{align*}
\]
Now, we first assume that Equations (14) and (15) are correct for the case with \( l = k \), that is,
\[
\pi_k^{(I)} = \left(\frac{\lambda}{k\theta} \delta_1^{d_1} \right) \frac{1}{d_2}
\]
and
\[
\pi_k^{(W)} = \delta_k.
\]
Then we need to check that Equations (14) and (15) are also correct for the case with \( l = k + 1 \). To that end, it follows from (14) that for \( l = k + 1 \)
\[
\lambda \delta_1^{d_1} - \lambda \delta_k^{d_1} + (k + 1) \theta \left[ \pi_{k+1}^{(I)} \right]^{d_2} + \lambda \left( \frac{\lambda}{k\theta} \delta_1^{d_1} \right) \frac{d_1}{d_2} - \mu \delta_k = 0.
\]
Noting that
\[
\delta_k = \rho \delta_1^{d_1} + \rho \left( \frac{\lambda}{k\theta} \delta_1^{d_1} \right) \frac{d_1}{d_2},
\]
we obtain
\[
\mu \delta_k = \lambda \delta_1^{d_1} + \lambda \left( \frac{\lambda}{k\theta} \delta_1^{d_1} \right) \frac{d_1}{d_2},
\]
hence we have
\[
\pi_{k+1}^{(I)} = \left[ \frac{\lambda}{(k + 1) \theta} \delta_1^{d_1} \right] \frac{1}{d_2},
\]
which, together with (15), follows that
\[
\pi_{k+1}^{(W)} = \frac{(k + 1) \theta}{\mu} \left[ \pi_{k+1}^{(I)} \right]^{d_2} + \rho \left[ \pi_{k+1}^{(I)} \right]^{d_1}
\]
\[=
\frac{(k + 1) \theta}{\mu} \cdot \lambda \delta_1^{d_1} + \rho \left[ \frac{\lambda}{(k + 1) \theta} \delta_1^{d_1} \right] \frac{d_1}{d_2}
\]
\[=
\rho \delta_k^{d_1} + \rho \left[ \frac{\lambda}{(k + 1) \theta} \delta_1^{d_1} \right] \frac{d_1}{d_2}
\]
\[=
\delta_{k+1}.
\]
This completes the proof. \( \blacksquare \)

The following corollary sets up a useful relation between \( \pi_k^{(I)} \) and \( \pi_k^{(W)} \) for \( k \geq 1 \).

**Corollary 2** For \( k \geq 1 \)
\[
k\theta \left[ \pi_k^{(I)} \right]^{d_2} = \lambda \left[ \pi_{k-1}^{(W)} \right]^{d_1},
\]
and there exists an integer \( K \geq 1 \) such that for \( k \geq K \)
\[
\left[ \pi_k^{(I)} \right]^{d_2} < \left[ \pi_{k-1}^{(W)} \right]^{d_1}.
\]
Furthermore, we have
\[
\lim_{k \to \infty} \frac{\left[\pi_k(I)\right]^{d_2}}{\left[\pi_k(W)\right]^{d_1}} = 0.
\]

**Proof:** It follows from (17) that for \( k \geq 1 \)
\[
\left[\pi_k(I)\right]^{d_2} = \frac{\lambda}{k \theta} \left[\pi_k(W)\right]^{d_1},
\]
which leads to
\[
k \theta \left[\pi_k(I)\right]^{d_2} = \lambda \left[\pi_k(W)\right]^{d_1}
\]
and
\[
\frac{\left[\pi_k(I)\right]^{d_2}}{\left[\pi_k(W)\right]^{d_1}} = \frac{\lambda}{k \theta}.
\]
Hence we have
\[
\lim_{k \to \infty} \frac{\left[\pi_k(I)\right]^{d_2}}{\left[\pi_k(W)\right]^{d_1}} = \lim_{k \to \infty} \frac{\lambda}{k \theta} = 0.
\]
Taking an integer \( K > \lambda / \theta \), we obtain that \( \left[\pi_k(I)\right]^{d_2} < \left[\pi_k(W)\right]^{d_1} \) for \( k \geq K \) due to \( \lambda / k \theta < 1 \). This completes the proof. \( \blacksquare \)

Using Theorem 1, we now consider a special case with \( d_1 = d_2 = d \geq 1 \). The following corollary provides an explicit expression for the fixed point with \( d_1 = d_2 = d \).

**Corollary 3** If \( d_1 = d_2 = d \geq 1 \), then the fixed point \( \pi = (\pi_0, \pi_1, \pi_2, ...) \) is given by
\[
\pi_k(I) = 1 - \eta_d,
\]
\[
\pi_k(W) = \rho (1 - \eta_d)^d,
\]
for \( k \geq 1 \)
\[
\pi_k(W) = \frac{\lambda + k \theta}{k \theta} \left[\frac{\lambda + (k - 1) \theta}{(k - 1) \theta}\right]^d \left[\frac{\lambda + (k - 2) \theta}{(k - 2) \theta}\right]^{d^2} \cdots \left(\frac{\lambda + \theta}{\theta}\right)^{d^{k-1}} \rho^{\frac{d^{k+1} - 1}{d-1}} (1 - \eta_d)^{d^{k+1}}
\]
and
\[
\pi_k(I) = \left(\frac{\lambda}{k \theta}\right)^d \frac{\lambda + (k - 1) \theta}{(k - 1) \theta} \left[\frac{\lambda + (k - 2) \theta}{(k - 2) \theta}\right]^{d^2} \cdots \left(\frac{\lambda + \theta}{\theta}\right)^{d^{k-2}} \rho^{\frac{d^{k+1} - 1}{d-1}} (1 - \eta_d)^{d^k}.
\]

16
Proof: By induction, we can easily derive the above desired result.

If \( d_1 = d_2 = d \geq 1 \), then

\[ \delta_0 = \rho (1 - \eta_d)^d \]

and for \( k \geq 1 \)

\[
\begin{align*}
\delta_k &= \rho \delta_{k-1} + \rho \delta_{k-1}^d \frac{\lambda}{k \theta} \\
&= \rho \left( \frac{\lambda}{k \theta} \delta_{k-1} \right) + \rho \delta_{k-1}^d \frac{\lambda}{k \theta} \\
&= \rho \frac{\lambda + k \theta}{k \theta} \delta_{k-1}^d \\
&= \rho \frac{\lambda + k \theta}{k \theta} \delta_{k-1}^d \\
&= \ldots \\
&= \frac{\lambda + k \theta}{k \theta} \left[ \frac{\lambda + (k - 1) \theta}{(k - 1) \theta} \right]^d \left[ \frac{\lambda + (k - 2) \theta}{(k - 2) \theta} \right]^d \ldots \left( \frac{\lambda + \theta}{\theta} \right)^{d^k} \rho \left( \frac{\lambda + k \theta}{k \theta} \delta_{k-1}^d \right) \\
&= \ldots \\
&= \frac{\lambda + k \theta}{k \theta} \left[ \frac{\lambda + (k - 1) \theta}{(k - 1) \theta} \right]^d \left[ \frac{\lambda + (k - 2) \theta}{(k - 2) \theta} \right]^d \ldots \left( \frac{\lambda + \theta}{\theta} \right)^{d^k} \rho \left( \frac{\lambda + k \theta}{k \theta} \delta_{k-1}^d \right) \\
\end{align*}
\]

This gives

\[
\begin{align*}
\pi_k^{(W)} &= \delta_k \\
&= \frac{\lambda + k \theta}{k \theta} \left[ \frac{\lambda + (k - 1) \theta}{(k - 1) \theta} \right]^d \left[ \frac{\lambda + (k - 2) \theta}{(k - 2) \theta} \right]^d \ldots \left( \frac{\lambda + \theta}{\theta} \right)^{d^k} \rho \left( \frac{\lambda + k \theta}{k \theta} \delta_{k-1}^d \right) \\
\end{align*}
\]

and

\[
\begin{align*}
\pi_k^{(I)} &= \left( \frac{\lambda}{k \theta} \theta_{k-1}^d \right)^{\frac{1}{d}} \\
&= \left( \frac{\lambda}{k \theta} \theta_{k-1}^d \right)^{\frac{1}{d}} \\
&= \left( \frac{\lambda}{k \theta} \theta_{k-1}^d \right)^{\frac{1}{d}} \\
&= \left( \frac{\lambda}{k \theta} \theta_{k-1}^d \right)^{\frac{1}{d}} \frac{\lambda + (k - 1) \theta}{(k - 1) \theta} \left[ \frac{\lambda + (k - 2) \theta}{(k - 2) \theta} \right]^d \ldots \left( \frac{\lambda + \theta}{\theta} \right)^{d^k} \rho \left( \frac{\lambda + k \theta}{k \theta} \delta_{k-1}^d \right) \\
\end{align*}
\]

This completes the proof.

Remark 2 It is easy to see from Theorem 2 that \( \pi_0^{(I)} \) and \( \pi_0^{(W)} \) are independent of the retrial processes including the two parameters \( \theta \) and \( d_2 \). Thus, the probing-server number \( d_2 \) will not have any influence on \( \pi_0^{(I)} \) and \( \pi_0^{(W)} \), but it can have influence on \( \pi_k^{(I)} \) and \( \pi_k^{(W)} \) for \( k \geq 1 \).

In the remainder of this section, numerical examples are used to illustrate how the fixed point depends on the probing-server numbers \( d_1 \) and/or \( d_2 \). In the examples, we take \( \lambda = 1, \mu = 5 \) and \( \theta = 2 \).
When \( d_2 = 1 \), Table 1 shows how the fixed point depends on the probing-server number \( d_1 \geq 2 \). It is easy to see from Table 1 that for a given \( d_1 \geq 2 \), \( \pi_k^{(I)} \) and \( \pi_k^{(W)} \) are two monotone decreasing functions of \( k \geq 1 \); and for a fixed \( k \geq 2 \), they decrease quickly to zero, as \( d_1 \) increases.

Table 1: The fixed point is a function of \( d_1 \) when \( d_2 = 1 \)

| \( d_1 \) | \( \pi_0^{(W)} \) | \( \pi_0^{(I)} \) | \( \pi_1^{(W)} \) | \( \pi_1^{(I)} \) | \( \pi_2^{(W)} \) | \( \pi_2^{(I)} \) |
|-----|------|------|------|------|------|------|
| 2   | 0.1459 | 0.8541 | 0.0043 | 0.0106 | 0 | 0 |
| 5   | 0.1110 | 0.8890 | 0 | 0 | 0 | 0 |
| 10  | 0.0836 | 0.9164 | 0 | 0 | 0 | 0 |

When \( d_1 = 1 \), Table 2 illustrates how \( \pi_k^{(W)} \) and \( \pi_k^{(I)} \), for a fixed \( k \geq 1 \), depend on the number \( d_2 \geq 2 \). It is seen that for a fixed \( k \geq 2 \), \( \pi_k^{(I)} \) and \( \pi_k^{(W)} \) are two monotone increasing functions of \( d_2 \geq 2 \). Also, the crucial number \( K_{d_1,d_2} \) given in Proposition 2 is shown as follows: \( K_{1,2} = K_{1,5} = 0 \), \( K_{1,8} = 1 \) and \( K_{1,10} = 5 \). Further, more numerical examples are used to illustrate that \( K_{1,d_2} \) is a monotone increasing function of \( d_2 \geq 2 \).

Table 2: The fixed point is a function of \( d_2 \) when \( d_1 = 1 \)

| \( d_2 \) | \( \pi_0^{(W)} \) | \( \pi_0^{(I)} \) | \( \pi_1^{(W)} \) | \( \pi_1^{(I)} \) | \( \pi_2^{(W)} \) | \( \pi_2^{(I)} \) | \( \pi_3^{(W)} \) | \( \pi_3^{(I)} \) |
|-----|------|------|------|------|------|------|------|------|
| 2   | 0.1667 | 0.8333 | 0.0911 | 0.2887 | 0.0484 | 0.1509 | 0.0276 | 0.0898 |
| 5   | 0.1667 | 0.8333 | 0.1550 | 0.6084 | 0.1354 | 0.5220 | 0.1208 | 0.4685 |
| 8   | 0.1667 | 0.8333 | 0.1799 | 0.7330 | 0.1717 | 0.6786 | 0.1626 | 0.6413 |
| 10  | 0.1667 | 0.8333 | 0.1893 | 0.7800 | 0.1853 | 0.7371 | 0.1783 | 0.7063 |
| 15  | 0.1667 | 0.8333 | 0.2028 | 0.8473 | 0.2045 | 0.8197 | 0.2006 | 0.7983 |
| 20  | 0.1667 | 0.8333 | 0.2100 | 0.8832 | 0.2146 | 0.8630 | 0.2122 | 0.8466 |

| \( d_2 \) | \( \pi_4^{(W)} \) | \( \pi_4^{(I)} \) | \( \pi_5^{(W)} \) | \( \pi_5^{(I)} \) | \( \pi_6^{(W)} \) | \( \pi_6^{(I)} \) |
|-----|------|------|------|------|------|------|
| 2   | 0.0173 | 0.0588 | 0.0118 | 0.0416 | 0.0086 | 0.0313 |
| 5   | 0.1106 | 0.4323 | 0.1034 | 0.4062 | 0.0980 | 0.3864 |
| 8   | 0.1554 | 0.6145 | 0.1499 | 0.5942 | 0.1456 | 0.5782 |
| 10  | 0.1724 | 0.6836 | 0.1677 | 0.6663 | 0.1640 | 0.6524 |
| 15  | 0.1965 | 0.7821 | 0.1932 | 0.7695 | 0.1905 | 0.7594 |
| 20  | 0.2093 | 0.8340 | 0.2067 | 0.8242 | 0.2046 | 0.8162 |
Table 3 illustrates that for $d_1 = 5$ and $a$ for a fixed $k \geq 1$, $\pi_k^{(I)}$ and $\pi_k^{(W)}$ are two monotone increasing functions of $d_2 \geq 2$. Since the number $d_1$ has a bigger impact on the fixed point than the number $d_2$, it is seen that $K_{5,d_2} = 0$ for $d_2 \geq 2$.

**Table 3: The fixed point is a function of $d_2$ when $d_1 = 5$**

| $d_2$ | $\pi_0^{(W)}$ | $\pi_0^{(I)}$ | $\pi_1^{(W)}$ | $\pi_1^{(I)}$ | $\pi_2^{(W)}$ | $\pi_2^{(I)}$ | $\pi_3^{(W)}$ | $\pi_3^{(I)}$ |
|-------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 5     | 0.1110         | 0.8890         | 0.0967         | 0              | 0              | 0              | 0              | 0              |
| 10    | 0.1110         | 0.8890         | 0.0006         | 0.3109         | 0              | 0.0210         | 0              | 0              |
| 15    | 0.1110         | 0.8890         | 0.0041         | 0.4589         | 0              | 0.1456         | 0              | 0.0209         |
| 20    | 0.1110         | 0.8890         | 0.0108         | 0.5576         | 0.0005         | 0.3007         | 0              | 0.1361         |
| 25    | 0.1110         | 0.8890         | 0.0193         | 0.6267         | 0.0029         | 0.4297         | 0.0004         | 0.2899         |
| 30    | 0.1110         | 0.8890         | 0.0285         | 0.6774         | 0.0082         | 0.5278         | 0.0027         | 0.4230         |

| $d_2$ | $\pi_4^{(W)}$ | $\pi_4^{(I)}$ | $\pi_5^{(W)}$ | $\pi_5^{(I)}$ | $\pi_6^{(W)}$ | $\pi_6^{(I)}$ |
|-------|----------------|----------------|----------------|----------------|----------------|----------------|
| 5     | 0              | 0              | 0              | 0              | 0              | 0              |
| 10    | 0              | 0              | 0              | 0              | 0              | 0              |
| 15    | 0              | 0.0008         | 0              | 0              | 0              | 0              |
| 20    | 0              | 0.0498         | 0              | 0.0140         | 0              | 0.0029         |
| 25    | 0.0001         | 0.1934         | 0              | 0.1278         | 0              | 0.0839         |
| 30    | 0.0010         | 0.3483         | 0.0004         | 0.2941         | 0.0002         | 0.2539         |

4 Expected Sojourn Time

In this section, we use the fixed point to provide an iterative algorithm for computing the expected sojourn time $E[T_{d_1,d_2}]$ that a tagged arriving customer spends in the retrial supermarket model.

If there are $k$ customers in the orbit, then we denote by $E\left[T^{(P)}_{d_1,d_2}(k)\right]$ and $E\left[T^{(R)}_{d_1,d_2}(k)\right]$ the expected sojourn times that a tagged primary arriving customer spends in the retrial supermarket model and that a tagged retrial arriving customer spends in the retrial supermarket model, respectively. At the same time, the probabilities that an arriving customer is primary and that an arriving customer is retrial are given by $\lambda/ (\lambda + k\theta)$ and
where we take $E$ between $E$ and $\frac{E}{\lambda + k\theta}$.

Note that $E$ finds one idle server and no customer is in the orbit, thus the expected sojourn time of the primary arriving customer is given by $1$ in this case, the tagged primary arriving customer has to enter the orbit so that this customer has the expected sojourn time $E$.

Using Equations (21) to (23), the following theorem provides an iterative algorithm for computing the expected sojourn time. The proof is easy and is omitted here.

**Theorem 4** For $k \geq 1$

$$E \left[ T_{d_1,d_2}^{(R)}(k) \right] = \frac{1}{\mu} \left\{ \frac{E}{\lambda} + \frac{k\theta}{\lambda + k\theta} \right\} + E \left[ T_{d_1,d_2}^{(P)}(k) \right] \right\},$$

(20)
Thus we obtain that for $d_1, d_2 \geq 1$.

Therefore, we have

$$E[T_{d_1, d_2} (0)] = \frac{1}{\mu} \left\{ \left[ \frac{\pi_0 (I)}{\pi_1 (I)} \right]^{d_1} - \left[ \frac{\pi_1 (I)}{\pi_2 (I)} \right]^{d_1} \right\}$$

$$+ \frac{1}{\mu} \left\{ \left[ \frac{\pi_0 (W)}{\pi_1 (W)} \right]^{d_1} - \left[ \frac{\pi_1 (W)}{\pi_2 (W)} \right]^{d_1} \right\} \left\{ \left[ \frac{\pi_1 (I)}{\pi_2 (I)} \right]^{d_2} - \left[ \frac{\pi_1 (W)}{\pi_2 (W)} \right]^{d_2} \right\}$$

and

$$E[T_{d_1, d_2} (k)] = \frac{1}{\mu} \left\{ \left[ \frac{\pi_k (I)}{\pi_{k+1} (I)} \right]^{d_1} - \left[ \frac{\pi_{k+1} (I)}{\pi_{k+2} (I)} \right]^{d_1} \right\}$$

$$+ \frac{1}{\mu} \left\{ \left[ \frac{\pi_k (W)}{\pi_{k+1} (W)} \right]^{d_1} - \left[ \frac{\pi_{k+1} (W)}{\pi_{k+2} (W)} \right]^{d_1} \right\} \left\{ \left[ \frac{\pi_{k+1} (I)}{\pi_{k+2} (I)} \right]^{d_2} - \left[ \frac{\pi_{k+1} (W)}{\pi_{k+2} (W)} \right]^{d_2} \right\}$$

Therefore, we have

$$E[T_{d_1, d_2}] = \sum_{k=0}^{\infty} \left\{ \frac{\lambda}{\lambda + k\theta} E[T_{d_1, d_2} (k)] + \frac{k\theta}{\lambda + k\theta} E[T_{d_1, d_2} (k)] \right\}.$$
Therefore, we get

\[
E \left[ T_{d_1,d_2} \right] = \sum_{k=0}^{\infty} \left\{ \frac{\lambda}{\lambda + k\theta} E \left[ T_{d_1,d_2}^{(P)} (k) \right] + \frac{k\theta}{\lambda + k\theta} E \left[ T_{d_1,d_2}^{(R)} (k) \right] \right\}
\]

\[
= \frac{1}{\mu} \left[ (1 - \eta d)^d - \frac{\lambda}{\theta} \delta_0^d \right] + \rho \frac{\{ \delta_1^d - \delta_2^d \} \{ \delta_2^d - \frac{1}{2} \theta \} \delta_1^d}{1 - \{ \delta_1^d - \delta_2^d \}}
\]

\[
+ \sum_{k=1}^{\infty} \frac{\lambda \rho}{\lambda + k\theta} \left\{ \frac{1}{k\theta} \delta_{k-1}^d - \frac{1}{(k+1)\theta} \delta_k^d \right\} \left\{ \frac{1}{\theta} \delta_{k+1}^d - \frac{1}{(k+2)\theta} \delta_{k+2}^d \right\}
\]

\[
+ \sum_{k=1}^{\infty} \frac{k\theta \rho}{\lambda + k\theta} \left\{ \frac{1}{k\theta} \delta_{k-1}^d - \frac{1}{(k+1)\theta} \delta_k^d \right\}.
\]

In the rest of this section, we provide numerical examples to show how the expected sojourn time \( E \left[ T_{d_1,d_2} \right] \) depends on the two probing-server numbers \( d_1 \) and \( d_2 \). In these numerical examples, we take that \( \lambda = 1, \mu = 5 \) and \( \theta = 2 \).

Figure 4 illustrates that the number \( d_1 \) has a bigger influence on the expected sojourn time than the number \( d_2 \). It can be observe that \( d_1 = 2 \) is a crucial point. If \( d_1 \geq 2 \), then influence of the number \( d_2 \) on the expected sojourn time is very weak because the three lines are almost identical. Figure 5 shows that the expected sojourn time is always monotone decreasing for \( d_2 \geq 1 \). It is seen that as the number \( d_1 \) increases, the expected sojourn time has a diminutively decreasing rate for \( d_2 \leq 100 \), but is can decrease quickly when \( d_2 \geq 1000 \). Figure 6 illustrates that the expected sojourn time is monotone decreasing for \( d_1 = d_2 = d \). It is seen from Figure 7 that the expected sojourn time is monotone decreasing for the two numbers \( d_1 \) and \( d_2 \), and the number \( d_1 \) plays a bigger role in the expected sojourn time than the number \( d_2 \).

**Remark 3** It is an interesting and difficult topic to study the role played by different probing-server numbers in performance analysis of supermarket models. Figures 4 to 7 and Tables 1 to 3 for our numerical examples give insights into the role of two probing-server numbers used by the primary and the retrial customers, respectively.

### 5 Exponential Convergence

In this section, we provide an upper bound for the current location of the retrial supermarket model. Based on this, we study exponential convergence of the current location.
Figure 4: The expected sojourn time depends on the number $d_1$

Figure 5: The expected sojourn time depends on the number $d_2$
Figure 6: The expected sojourn time depends on the numbers $d_1 = d_2 = d$.

Figure 7: The expected sojourn time depends on the numbers $d_1, d_2$. 
to the fixed point.

For the retrial supermarket model, the initial point $S(0)$ can affect the current location $S(t)$ for each $t > 0$, since the arrival, service and retrial processes in the retrial supermarket model are under a unified structure through a sample path comparison. To explain this, it is necessary to provide some notation for comparison of two vectors. Let $a = (a_1, a_2, a_3, \ldots)$ and $b = (b_1, b_2, b_3, \ldots)$. We write $a < b$ if $a_k < b_k$ for some $k \geq 1$ and $a_l \leq b_l$ for $l \neq k, l \geq 1$; and $a \leq b$ if $a_k \leq b_k$ for all $k \geq 1$.

Now, we can easily obtain the following proposition, while the proof is clear from a sample path analysis and is omitted here.

**Proposition 4** If $S(0) \preceq \tilde{S}(0)$, then $S(t) \preceq \tilde{S}(t)$.

Based on Proposition 4, the following theorem shows that the fixed point $\pi$ is an upper bound of the current location $S(t)$ for all $t \geq 0$.

**Theorem 5** For the retrial supermarket model, if there exists some $k$ such that $S_k(I)(0) = 0$ or $S_k(W)(0) = 0$, then the current location $S(t)$ for all $t \geq 0$ has an upper bound vector $\pi$, that is, $S(t) \preceq \pi$ for all $t \geq 0$.

**Proof:** Let
\[
\tilde{S}_k(I)(0) = \pi_k(I), \tilde{S}_k(W)(0) = \pi_k(W), \quad k \geq 0.
\]
Then for each $k \geq 0$, $\tilde{S}_k(I)(t) = \tilde{S}_k(I)(0) = \pi_k(I)$ and $\tilde{S}_k(W)(t) = \tilde{S}_k(W)(0) = \pi_k(W)$ for all $t \geq 0$, since $\tilde{S}(0) = \pi$ is the fixed point for the retrial supermarket model. If $S_k(I)(0) = 0$ or $S_k(W)(0) = 0$ for some $k$, then $S_k(I)(0) \prec \tilde{S}_k(I)(0)$ or $S_k(W)(0) \prec \tilde{S}_k(W)(0)$. Thus, we have $S(0) \preceq \tilde{S}(0)$. It is easy to see from Proposition 4 that $S_k(I)(t) \preceq \tilde{S}_k(I)(t) = \pi_k(I)$ and $S_k(W)(t) \preceq \tilde{S}_k(W)(t) = \pi_k(W)$ for all $k \geq 0$ and $t \geq 0$, which lead to $S(t) \preceq \pi$ for all $t \geq 0$. This completes the proof.

To show exponential convergence, we apply a potential function (or Lyapunov function) $\Phi(t)$ which is defined as
\[
\Phi(t) = \sum_{k=0}^{\infty} w_k \left[ \pi_k(I) - S_k(I)(t) \right] + \sum_{k=0}^{\infty} v_k \left[ \pi_k(W) - S_k(W)(t) \right],
\]
where $\{w_k\}$ and $\{v_k\}$ are two underdetermined constant sequences with $w_0 = 1$.

The following theorem measures the distance of the current location $S(t)$ for $t \geq 0$ to the fixed point $\pi$, and illustrates that the $L_1$ distance to the fixed point from the current
location is very close to zero with exponential convergence. Hence, it shows that from any suitable starting point, the retrial supermarket model can quickly converge to the fixed point, that is, there always exists a fixed point in the retrial supermarket model.

**Theorem 6** For \( t \geq 0 \),

\[
\Phi(t) \leq c_0 e^{-\delta t},
\]

where \( c_0 \) and \( \delta \) are two positive constants. In this case, the potential function \( \Phi(t) \) is exponentially convergent.

**Proof:** It is seen from (24) that

\[
\Phi(t) = \sum_{k=0}^{\infty} w_k \left[ \pi_k^{(I)} - S_k^{(I)}(t) \right] + \sum_{k=0}^{\infty} v_k \left[ \pi_k^{(W)} - S_k^{(W)}(t) \right],
\]

which leads to

\[
\frac{d}{dt} \Phi(t) = -\sum_{k=0}^{\infty} w_k \frac{d}{dt} S_k^{(I)}(t) - \sum_{k=0}^{\infty} v_k \frac{d}{dt} S_k^{(W)}(t).
\]

It follows from Equations (7) to (10) that

\[
-\sum_{k=0}^{\infty} w_k \frac{d}{dt} S_k^{(I)}(t) = \sum_{k=0}^{\infty} \lambda w_k \left[ S_k^{(I)}(t) \right] \frac{d_1}{d} - \sum_{k=1}^{\infty} k\theta w_k \left[ S_k^{(I)}(t) \right] \frac{d_2}{d} - \mu w_k S_k^{(W)}(t)
\]

and

\[
-\sum_{k=0}^{\infty} v_k \frac{d}{dt} S_k^{(W)}(t) = -\sum_{k=0}^{\infty} \lambda v_k \left[ S_k^{(W)}(t) \right] \frac{d_1}{d} - \sum_{k=1}^{\infty} k\theta v_{k-1} \left[ S_k^{(I)}(t) \right] \frac{d_2}{d} + \sum_{k=0}^{\infty} \lambda \left( v_k - v_{k+1} \right) \left[ S_k^{(W)}(t) \right] \frac{d_1}{d} + \sum_{k=0}^{\infty} \mu v_k S_k^{(W)}(t),
\]

Thus, we obtain

\[
\frac{d}{dt} \Phi(t) = \sum_{k=0}^{\infty} \lambda \left( w_k - v_k \right) \left[ S_k^{(I)}(t) \right] \frac{d_1}{d} + \sum_{k=1}^{\infty} \lambda \left( w_k - v_{k-1} \right) \left[ S_k^{(I)}(t) \right] \frac{d_2}{d} + \sum_{k=0}^{\infty} \lambda \left( v_k - v_{k+1} \right) \left[ S_k^{(W)}(t) \right] \frac{d_1}{d} + \sum_{k=0}^{\infty} \mu \left( v_k - w_k \right) S_k^{(W)}(t).
\]

Let

\[
\left[ S_k^{(I)}(t) \right] \frac{d_1}{d} = c_k(t) \left[ \pi_k^{(I)} - S_k^{(I)}(t) \right],
\]

\[
\left[ S_k^{(I)}(t) \right] \frac{d_2}{d} = d_k(t) \left[ \pi_k^{(I)} - S_k^{(I)}(t) \right],
\]

26
\[
\left[ S_k^{(W)} (t) \right]_{d_1} = g_k (t) \left[ \pi_k^{(W)} - S_k^{(W)} (t) \right]
\]

and
\[
S_k^{(W)} (t) = h_k (t) \left[ \pi_k^{(W)} - S_k^{(W)} (t) \right]
\]

Then
\[
\frac{d}{dt} \Phi (t) = \sum_{k=0}^{\infty} \lambda (w_k - v_k) c_k (t) \left[ \pi_k - S_k^{(I)} (t) \right] + \sum_{k=1}^{\infty} k\theta (w_k - v_{k-1}) d_k (t) \left[ \pi_k - S_k^{(I)} (t) \right] + \sum_{k=0}^{\infty} \lambda (v_k - v_{k+1}) g_k (t) \left[ \pi_k - S_k^{(W)} (t) \right] + \sum_{k=0}^{\infty} \mu (v_k - w_k) h_k (t) \left[ \pi_k - S_k^{(W)} (t) \right] = \lambda (w_0 - v_0) c_0 (t) \left[ \pi_0 - S_0^{(I)} (t) \right] + \sum_{k=1}^{\infty} \lambda (w_k - v_k) c_k (t) + k\theta (w_k - v_{k-1}) d_k (t) \left[ \pi_k - S_k^{(I)} (t) \right] + \sum_{k=0}^{\infty} \lambda (v_k - v_{k+1}) g_k (t) + \mu (v_k - w_k) h_k (t) \left[ \pi_k - S_k^{(W)} (t) \right].
\]

Let
\[
w_0 = 1,
\]
\[
\lambda (w_0 - v_0) c_0 (t) \leq -\delta v_0,
\]
for \( k \geq 1 \)
\[
\lambda (w_k - v_k) c_k (t) + k\theta (w_k - v_{k-1}) d_k (t) \leq -\delta v_k;
\]
and for \( l \geq 0 \)
\[
\lambda (v_l - v_{l+1}) g_l (t) + \mu (v_l - w_l) h_l (t) \leq -\delta w_l.
\]

Then
\[
\frac{d}{dt} \Phi (t) \leq -\delta \left\{ \sum_{k=0}^{\infty} w_k \left[ \pi_k^{(I)} - S_k^{(I)} (t) \right] + \sum_{k=0}^{\infty} v_k \left[ \pi_k^{(W)} - S_k^{(W)} (t) \right] \right\}.
\]

This gives
\[
\frac{d}{dt} \Phi (t) \leq -\delta \Phi (t) ,
\]
thus we obtain
\[
\Phi (t) \leq c_0 e^{-\delta t}.
\]
This completes the proof. 

**Remark 4** We provide an algorithm for computing the two underdetermined constant sequences \( \{w_k\} \) and \( \{v_k\} \) with \( w_0 = 1 \) as follows:

**Step one:**

\[
    w_0 = 1.
\]

**Step two:**

\[
    v_0 = \frac{\lambda}{\lambda G_0(t) - \delta}.
\]

**Step three:**

\[
    v_1 = \frac{\delta + \lambda v_0 g_0(t) + \mu (v_0 - 1) h_0(t)}{\lambda g_0(t)}
\]

and

\[
    w_1 = \frac{\lambda v_1 c_1(t) + \theta v_0 d_1(t) - \delta v_1}{\lambda c_1(t) + \theta d_1(t)}.
\]

**Step four:** for \( k \geq 2 \)

\[
    v_k = \frac{\delta w_{k-1} + \lambda v_{k-1} g_{k-1}(t) + \mu (v_{k-1} - w_{k-1}) h_{k-1}(t)}{\lambda g_{k-1}(t)}
\]

and

\[
    w_k = \frac{\lambda v_k c_k(t) + k \theta v_{k-1} d_k(t) - \delta v_k}{\lambda c_k(t) + k \theta d_k(t)}.
\]

6 A Lipschitz Condition

In this section, we apply the Kurtz Theorem to study the density-dependent jump Markov process corresponding to the retrial supermarket model with Poisson arrivals, exponential service times and exponential retrial times and with two different probing-server numbers. This will lead to the Lipschitz condition under which the fraction measure of the retrial supermarket model weakly converges to the system of differential equations.

The retrial supermarket model can be analyzed by a density-dependent jump Markov process, which is a Markov process with the population size \( n \). Kurtz’s work provides a basis for density-dependent jump Markov processes in order to relate infinite-size systems of differential equations to corresponding finite-size systems of differential equations. Readers may refer to Kurtz [16] for more details.

In the retrial supermarket model with \( n \) servers, we denote by \( n_k^{(W)} \) the numbers of busy servers with at least \( k \) customers in the orbit, and \( n_k^{(I)} \) the numbers of idle servers
with at least \( k \) customers in the orbit. Hence \( 0 \leq n_k(W), n_k(I) \leq n \) for \( k \geq 0 \). Note that \( n_k(W) \) and \( n_k(I) \) are independent of time \( t \geq 0 \), thus they are different from \( n_k(W)(t) \) and \( n_k(I)(t) \) defined in Section 2 for the retrial supermarket models.

We write
\[
E'_n(W) = \{(W, k) : 0 \leq k \leq n\}
\]
and
\[
E'_n(I) = \{(I, k) : 0 \leq k \leq n\},
\]
\[
E_n = E_n(W) \cup E_n(I).
\]

In the state space \( E_n \), the density-dependent jump Markov process \( \{X_n(t) : t \geq 0\} \), defined in Section 2, for the retrial supermarket model contains four classes of state transitions:

Class one: \( E_n(W) \overset{a}{\rightarrow} E_n(W): (W, i) \rightarrow (W, i + 1) \),

Class two: \( E_n(W) \overset{s}{\rightarrow} E_n(I): (W, i) \rightarrow (I, i) \),

Class three: \( E_n(I) \overset{a}{\rightarrow} E_n(W): (I, i) \rightarrow (W, i) \), and

Class four: \( E_n(I) \overset{r}{\rightarrow} E_n(W): (I, i) \rightarrow (W, i - 1) \).

Note that the state transitions \( a \), \( s \) and \( r \) express a primary arrival, a service completion and a retrial arrival, respectively.

For \( k \geq 0 \) we write
\[
s_k(I)(n) = E\left[\frac{n_k(I)}{n}\right]
\]
and
\[
s_k(W)(n) = E\left[\frac{n_k(W)}{n}\right].
\]

Let
\[
S_0(n) = s^{(I)}_0(n)
\]
and for \( k \geq 1 \)
\[
S_k(n) = \left(s^{(W)}_{k-1}(n), s^{(I)}_k(n)\right),
\]
\[
S(n) = (S_0(n), S_1(n), S_2(n), \ldots).
\]

Note that the states of the density-dependent jump Markov process \( \{X_n(t) : t \geq 0\} \) can be normalized and be interpreted as measuring population densities: \( S(n) = (S_0(n), S_1(n), S_2(n), \ldots) \).

The transition rates of this Markov process also depend on these population densities.
Let \( \{ \hat{X}_n (t) : t \geq 0 \} \) be a density-dependent jump Markov process on the state space \( E_n \) whose transition rates corresponding to the above four cases of state transitions are given by

\[
q^{(n)}_{(W,k) \to (W,k+1)} = n \beta_k a_{k+1} \left( s_k^{(W)} (n) \right) = n \lambda s_k^{(W)} (n), \tag{25}
\]

\[
q^{(n)}_{(W,k) \to (I,k)} = n \beta_k s_k \left( s_k^{(I)} (n) \right) = n \mu s_k^{(I)} (n), \tag{26}
\]

\[
q^{(n)}_{(I,k) \to (W,k)} = n \beta_k a_{k} \left( s_k^{(I)} (n) \right) = n \lambda s_k^{(I)} (n), \tag{27}
\]

\[
q^{(n)}_{(I,k) \to (W,k-1)} = n \beta_k r_{k-1} \left( s_k^{(I)} (n) \right) = nk \theta s_k^{(I)} (n). \tag{28}
\]

In the retrial supermarket model, \( \hat{X}_n (t) \) is an unscaled process which records either the number of busy servers or the number of idle servers with at least \( k \) customers in the orbit for \( k \geq 0 \). Using Chapter 7 in Kurtz [16] or Subsection 3.4.1 in Mitzenmacher [28], the Markov process \( \{ \hat{X}_n (t) : t \geq 0 \} \) with the transition rates, described in Equations (25) to (28), is given by

\[
\hat{X}_n (t) = \hat{X}_n (0) + \sum_{b \in E} l_b Y_b \left( n \int_0^t \beta_b \left( \frac{\hat{X}_n (u)}{n} \right) du \right), \tag{29}
\]

where \( Y_b (x) \) are independent standard Poisson processes and \( l_b \) is a positive integer with \( l_b \leq \Re < +\infty \) for \( b \in E \), where

\[
E = \{(W,k) \overset{a}{\rightarrow} (W,k+1), (W,k) \overset{s}{\rightarrow} (I,k), (I,k) \overset{a}{\rightarrow} (W,k); (I,k+1) \overset{r}{\rightarrow} (W,k) : k \geq 0\}.
\]

Clearly, the state of the density-dependent jump Markov process in Equation (29) at time \( t \) is determined by the starting point and the transition rates in Equations (25) to (28), which are integrated over its history.

Let

\[
F (y) = \sum_{b \in E (y)} l_b \beta_b (y), \tag{30}
\]

where

\[
E (y) = \{ b \in E : \text{A state transition } b \text{ is directly related to state } y \}.
\]

That is, the state transition \( b \) from state \( y \) contain two classes: 1) \( \overset{b}{\rightarrow} \) for entering state \( y \), and 2) \( \overset{b}{\rightarrow} \) for leaving from state \( y \). Taking \( X_n (t) = n^{-1} \hat{X}_n (t) \) which is an appropriate scaled process, we have

\[
X_n (t) = X_n (0) + \sum_{b \in E} l_b n^{-1} \hat{Y}_b \left( n \int_0^t \beta_b (X_n (u)) du \right) + \int_0^t F (X_n (u)) du, \tag{31}
\]

30
where $\hat{Y}_b(y) = Y_b(y) - y$ is a Poisson process centered at its expectation. Let $X(t) = \lim_{n \to \infty} X_n(t)$ and $x_0 = \lim_{n \to \infty} X_n(0)$, we obtain

$$X(t) = x_0 + \int_0^t F(X(u)) \, du, \quad t \geq 0, \quad (32)$$

due to the fact that

$$\lim_{n \to \infty} \frac{1}{n} \hat{Y}_b(n \int_0^t \beta_b(X_n(u)) \, du) = 0$$

by means of the law of large numbers. In the retrial supermarket model, the deterministic and continuous process $\{X(t), t \geq 0\}$ is described by the infinite-size system of differential equations (7) to (10), or simply,

$$\frac{d}{dt}X(t) = F(X(t)) \quad (33)$$

with the initial condition

$$X(0) = x_0. \quad (34)$$

In the rest of this section, it is necessary to express the function $F(y)$ in order to set up a Lipschitz condition. To that end, using the law of large numbers we have

$$\pi_k = \lim_{n \to \infty} S_k(n), \quad k \geq 0.$$ 

Let

$$\Omega = \left\{ \pi_0^{(I)}, \pi_1^{(I)}, \pi_0^{(W)}, \pi_2^{(I)}, \pi_1^{(W)}, \ldots \right\}.$$ 

Then it is easy to see from the system of differential equations (7) to (10) that the function $F(y)$ is given by

$$F(y) = \begin{cases} 
-\lambda \left[ \pi_0^{(I)} \right] d_1 + \mu \pi_0^{(W)}, & \text{if } y = \pi_0^{(I)}, \\
-\lambda \left[ \pi_k^{(I)} \right] d_1 - k\theta \left[ \pi_k^{(I)} \right] d_2 + \mu \pi_k^{(W)}, & \text{if } y = \pi_k^{(I)} \text{ for } k \geq 1, \\
\lambda \left[ \pi_0^{(W)} \right] d_1 - \lambda \left[ \pi_k^{(W)} \right] d_1 + \theta \left[ \pi_k^{(I)} \right] d_2 - \mu \pi_0^{(W)}, & \text{if } y = \pi_0^{(W)}, \\
\lambda \left[ \pi_{k-1}^{(W)} \right] d_1 - \lambda \left[ \pi_k^{(W)} \right] d_1 + (k + 1)\theta \left[ \pi_{k+1}^{(I)} \right] d_2 - \mu \pi_k^{(W)}, & \text{if } y = \pi_k^{(W)} \text{ for } k \geq 1.
\end{cases} \quad (35)$$

Now, we consider uniqueness of the limiting deterministic process $\{X(t), t \geq 0\}$ with (33) and (34), or uniqueness of solution to the infinite-size system of differential equations (7) to (10). To that end, a sufficient condition for uniqueness of solution is Lipschitz, that is, there exists a number $M > 0$ such that for $y, z \in \Omega$

$$|F(y) - F(z)| \leq M||y - z||.$$
In general, the Lipschitz condition is standard and sufficient for uniqueness of solution to the finite-size system of differential equations; while for the countable infinite-size case, readers may refer to Theorem 3.2 in Deimling [7] and Subsection 3.4.1 in Mitzenmacher [28] for some useful generalization of the Lipschitz condition.

The following theorem shows that the retrial supermarket model satisfies the Lipschitz condition for analyzing uniqueness of solution to the infinite-size system of differential equations (7) to (10).

**Theorem 7** The Lipschitz condition is satisfied in the retrial supermarket model with Poisson arrivals, exponential service times and exponential retrial times and with two different probing-server numbers.

**Proof** For the retrial supermarket model, it is easy to see from (35) that we need to consider eight types of pairs \((y, z)\) in order to check that there exists a bigger \(M > 0\) such that \(|F(y) - F(z)| \leq M ||y - z||\) for any \(y, z \in \Omega\). In what follows we only prove two types of pairs \((y, z)\), while all the others can be proved similarly.

Let \(y = \pi_0^{(I)}\) and \(z = \pi_k^{(I)}\) for \(k \geq 1\). Then

\[
|F(y) - F(z)| \leq \lambda \left\{ \left[ \pi_0^{(I)} \right]^{d_1} - \left[ \pi_k^{(I)} \right]^{d_1} + \mu \left| \pi_0^{(W)} - \pi_k^{(W)} \right| + k\theta \left[ \pi_k^{(I)} \right]^{d_2} \right. \]

Note that

\[
\pi_k^{(I)} = \left( \frac{\lambda}{k\theta d_{k-1}} \right)^{\frac{1}{d_2}}
\]

which leads to

\[
k\theta \left[ \pi_k^{(I)} \right]^{d_2} = \lambda d_{k-1}.
\]

Since \(\pi_0^{(I)} > \pi_k^{(I)}\) and \(\pi_0^{(W)} > \pi_k^{(W)}\), we have

\[
|F(y) - F(z)| \leq \lambda \left\{ \left[ \pi_0^{(I)} \right]^{d_1} - \left[ \pi_k^{(I)} \right]^{d_1} + \mu \left| \pi_0^{(W)} - \pi_k^{(W)} \right| + \lambda d_{k-1} \right. \]

\[
< \lambda \left[ \pi_0^{(I)} \right]^{d_1} + \mu \pi_0^{(W)} + \lambda \left[ \pi_k^{(W)} \right]^{d_1}
\]

\[
< \lambda \left[ \pi_0^{(I)} \right]^{d_1} + \mu \pi_0^{(W)} + \lambda \left[ \pi_0^{(W)} \right]^{d_1}
\]

due to the fact that \(\pi_0^{(W)} > \pi_k^{(W)} > \pi_k^{(W)}\) for \(k \geq 2\). Since \(\pi_0^{(I)} > \pi_k^{(I)} > \pi_k^{(I)}\) for \(k \geq 2\), we have

\[
\pi_0^{(I)} - \pi_1^{(I)} < \pi_0^{(I)} - \pi_k^{(I)}.
\]
which leads to
\[
\lambda \left[ \frac{\pi_0(I)}{\pi_0(I) - \pi_1(I)} \right]^{d_1} + \mu \pi_0(W) + \lambda \left[ \frac{\pi_0(W)}{\pi_0(I) - \pi_1(I)} \right]^{d_1} > \lambda \left[ \frac{\pi_0(I)}{\pi_0(I) - \pi_1(I)} \right]^{d_1} + \mu \pi_0(W) + \lambda \left[ \frac{\pi_0(W)}{\pi_0(I) - \pi_1(I)} \right]^{d_1}.
\]

Taking
\[
M_1 = \frac{\lambda \left[ \frac{\pi_0(I)}{\pi_0(I) - \pi_1(I)} \right]^{d_1} + \mu \pi_0(W) + \lambda \left[ \frac{\pi_0(W)}{\pi_0(I) - \pi_1(I)} \right]^{d_1}}{\pi_0(I) - \pi_1(I)},
\]
we obtain
\[
|F(y) - F(z)| \leq M_1 \left[ \frac{\pi_0(I) - \pi_1(I)}{\pi_0(I) - \pi_1(I)} \right] = M_1 |y - z|.
\]

Let \( y = \pi_1(W) \) and \( z = \pi_k(W) \) for \( k \geq 2 \). Then
\[
|F(y) - F(z)| \leq \lambda \left[ \frac{\pi_1(W)}{\pi_1(W) - \pi_2(W)} \right]^{d_1} + \lambda \left[ \frac{\pi_1(W)}{\pi_1(W) - \pi_2(W)} \right]^{d_1} | + \lambda \left[ \frac{\pi_1(W)}{\pi_1(W) - \pi_2(W)} \right]^{d_1}
\]
\[
+ 2\theta \left[ \frac{\pi_2(W)}{\pi_2(W) - (k + 1)} \right]^{d_2} - (k + 1) \theta \left[ \frac{\pi_2(W)}{\pi_2(W) - (k + 1)} \right]^{d_2} + \mu \pi_1(W) - \pi_k(W) + \lambda \left[ \frac{\pi_1(W)}{\pi_1(W) - \pi_2(W)} \right]^{d_1} - \left[ \frac{\pi_k(W)}{\pi_k(W)} \right]^{d_1}|
\]
\[
\leq 2\lambda \left[ \frac{\pi_1(W)}{\pi_1(W) - \pi_2(W)} \right]^{d_1} + 2\theta \left[ \frac{\pi_2(W)}{\pi_2(W) - (k + 1)} \right]^{d_2} + \lambda \left[ \frac{\pi_1(W)}{\pi_1(W) - \pi_2(W)} \right]^{d_1} + \mu \pi_1(W)
\]
\[
= 3\lambda \left[ \frac{\pi_1(W)}{\pi_1(W) - \pi_2(W)} \right]^{d_1} + \lambda \left[ \frac{\pi_1(I)}{\pi_1(I) - \pi_2(I)} \right]^{d_1} + \mu \pi_1(W).
\]

Note that for \( k \geq 2 \)
\[
\pi_1(W) - \pi_2(W) \leq \pi_1(W) - \pi_k(W),
\]
which follows
\[
3\lambda \left[ \frac{\pi_1(W)}{\pi_1(W) - \pi_2(W)} \right]^{d_1} + \lambda \left[ \frac{\pi_1(I)}{\pi_1(I) - \pi_2(I)} \right]^{d_1} + \mu \pi_1(W) \geq 3\lambda \left[ \frac{\pi_1(W)}{\pi_1(W) - \pi_2(W)} \right]^{d_1} + \lambda \left[ \frac{\pi_1(I)}{\pi_1(I) - \pi_2(I)} \right]^{d_1} + \mu \pi_1(W)
\]
\[
= M_2 \left[ \frac{\pi_1(W) - \pi_k(W)}{\pi_1(W) - \pi_k(W)} \right] = M_2 |y - z|.
\]

Let
\[
M_2 = \frac{3\lambda \left[ \frac{\pi_1(W)}{\pi_1(W) - \pi_2(W)} \right]^{d_1} + \lambda \left[ \frac{\pi_1(I)}{\pi_1(I) - \pi_2(I)} \right]^{d_1} + \mu \pi_1(W)}{\pi_1(W) - \pi_2(W)}.
\]

Then
\[
|F(y) - F(z)| \leq M_2 \left[ \frac{\pi_1(W) - \pi_k(W)}{\pi_1(W) - \pi_k(W)} \right] = M_2 |y - z|.
\]

Similar to the above analysis, for the ith type of pairs \((y, z)\) there exists a number \(M_i > 0\) such that
\[
|F(y) - F(z)| \leq M_i |y - z|.
\]

Taking \( M = \max \{M_i : i = 1, 2, \ldots, 8\} \), we obtain that for any \( y, z \in \Omega \),
\[
|F(y) - F(z)| \leq M |y - z|.
\]
This completes the proof.

Based on Theorem 7, the following theorem directly follows from Theorem 3.13 in Mitzenmacher [28].

**Theorem 8** In the retrial supermarket model with Poisson arrivals, exponential service times and exponential retrial times and with two different probing-server numbers, we have

\[
\lim_{n \to \infty} \sup_{u \leq t} |X_n(u) - X(u)| = 0, \quad \text{a.s.},
\]

where \( \{X_n(t)\} \) and \( \{X(t)\} \) are given by (31) and (32), respectively.

**Proof** For the retrial supermarket model, Theorem 7 shows that the function \( F(y) \) for \( y \in \Omega \) satisfies the Lipschitz condition. Based on this, it is easy to take a subset \( \Omega^* \subset \Omega \) such that

\[
\{X(u) : u \leq t\} \subset \Omega^*
\]

and

\[
\sup_{y \in \Omega^*} F(y) < +\infty.
\]

Thus, this proof can be further completed by means of Theorem 3.13 in Mitzenmacher [28]. This completes the proof.

Using Theorem 3.11 in Mitzenmacher [28] and Theorem 8, the following theorem provides an upper bound over the time interval \([0, t]\) for the expected sojourn time that an arriving tagged customer spends in an initially empty retrial supermarket model.

**Theorem 9** For the retrial supermarket model with Poisson arrivals, exponential service times and exponential retrial times and with two different probing-server numbers, the expected sojourn time that a tagged arriving customer spends in an initially empty system over the time interval \([0, t]\) is bounded above by

\[
\sum_{k=0}^{\infty} \left\{ \frac{\lambda}{\lambda + k\theta} E\left[T_S^{(P)}(k)\right] + \frac{k\theta}{\lambda + k\theta} E\left[T_S^{(R)}(k)\right] \right\} + o(1),
\]

where \( o(1) \) is understood from the population size \( n \to \infty \).

7 Concluding remarks

In this paper, we have presented a novel approach for studying super-exponential solution of the retrial supermarket model with Poisson arrivals, exponential service times and
exponential retrial times, and with two different probing-server numbers. We describe the retrial supermarket model as a system of differential equations by means of density-dependent jump Markov processes, and obtain an iterative algorithm for computing the fixed point of the system of differential equations. Based on the fixed point, we analyze the expected sojourn time that a tagged arriving customer spends in this system, and use numerical examples to indicate different influence of the two probing-server numbers on system performance including the expected sojourn time. Furthermore, we analyze exponential convergence of the current location of the retrial supermarket model to the fixed point, and apply the Kurtz Theorem to study density-dependent jump Markov process given in the retrial supermarket model. This leads to a Lipschitz condition under which the fraction measure of the retrial supermarket model weakly converges to the system of differential equations. This paper gives a new understanding of how the workload probing can help in load balancing jobs in retrial supermarket models.

Our approach given in this paper is useful in the study of retrial supermarket models with applications to, such as data centers, multi-core servers systems and cloud computational modeling. We expect that this approach will be applicable to the study of more general retrial supermarket models, such as non-Poisson arrival processes, non-exponential service times, and retrial supermarket networks (for example, each server contains $c$ identical processors). It will also open a new avenue to quantitative evaluation of more general supermarket models and analysis of retrial queueing networks.

Acknowledgements

The authors are grateful to Professor Henk C. Tijms whose comments greatly help us to improve the presentation of this paper. Q.L. Li was supported by the National Science Foundation of China under grant No. 10871114, John C.S. Lui was supported by the RGC grant, and Yang Wang was supported by the National Science Foundation of China under grant No. 61001075.

References

[1] J.R. Artalejo (1999). A classified bibliography of research on retrial queues: Progress in 1990-1999. *Top 7*, 187–211.
[2] J.R. Artalejo and A. Gómez-Corral (2008). *Retrial Queueing Systems: A Computational Approach*. Springer.

[3] Y. Azar, A.Z. Broder, A.R. Karlin and E. Upfal (1999). Balanced allocations. *SIAM Journal on Computing* **29**, 180–200.

[4] M. Bramson, Y. Lu and B. Prabhakar (2010). Randomized load balancing with general service time distributions. In *Proceedings of the ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, pages 275–286.

[5] M. Dahlin (1999). Interpreting stale load information. *IEEE Transactions on Parallel and Distributed Systems* **11**, 1033–1047.

[6] R.W.R. Darling and J.R. Norris (2008). Differential equation approximations for Markov chains. *Probability Surveys* **5**, 37–79.

[7] K. Deimling (1977). *Ordinary Differential Equations in Banach Spaces*. Springer-Verlag.

[8] G.I. Falin (1990). A survey of retrial queues. *Queueing Systems* **7**, 127-167.

[9] G.I. Falin and J.G.C. Templeton (1997). *Retrial Queues*. Chapman & Hall.

[10] A. Gómez-Corral (2006). A bibliographical guide to the analysis of retrial queues through matrix analytic techniques. *Annals of Operations Research* **141**, 163–191.

[11] C. Graham (2000). Kinetic limits for large communication networks. In *Modelling in Applied Sciences*, N. Bellomo and M. Pulvirenti (eds.), Birkhäuser, pages 317–370.

[12] C. Graham (2000). Chaoticity on path space for a queueing network with selection of the shortest queue among several. *Journal of Applied Probability* **37**, 198–201.

[13] C. Graham (2004). Functional central limit theorems for a large network in which customers join the shortest of several queues. *Probability Theory Related Fields* **131**, 97–120.

[14] M. Harchol-Balter and A.B. Downey (1997). Exploiting process lifetime distributions for dynamic load balancing. *ACM Transactions on Computer Systems* **15**, 253–285.
[15] V.G. Kulkarni and H.M. Liang (1997). Retrial queues revisited. In *Frontiers in Queueing: Models and Applications in Science and Engineering*, J.H. Dshalalow (ed), CRC Press, pages 19-34.

[16] T.G. Kurtz (1981). *Approximation of Population Processes*. SIAM.

[17] Q.L. Li (2010). *Constructive Computation in Stochastic Models with Applications: The RG-Factorizations*. Springer and Tsinghua Press.

[18] Q.L. Li (2010). Doubly exponential solution for randomized load balancing with general service times. Submitted for publication.

[19] Q.L. Li and John C.S. Lui (2010). Doubly exponential solution for randomized load balancing models with Markovian arrival processes and PH service times. Submitted for publication.

[20] Q.L. Li, John C.S. Lui and Y. Wang (2010). A matrix-analytic solution for randomized load balancing models with PH service times. In *Proceeding of PERFORM 2010 Workshop*, Lecture Notes of Computer Science, Pages 1-20.

[21] Q.L. Li, John C.S. Lui and Y. Wang (2010). An iterative matrix-analytic algorithm for supermarket models with batch Markovian arrival processes. Submitted for publication.

[22] M. Luczak and C. McDiarmid (2006). On the maximum queue length in the supermarket model. *The Annals of Probability* **34**, 493–527.

[23] M. Luczak and C. McDiarmid (2007). Asymptotic distributions and chaos for the supermarket model. *Electronic Journal of Probability* **12**, 75–99.

[24] M.J. Luczak and J.R. Norris (2005). Strong approximation for the supermarket model. *The Annals of Applied Probability* **15**, 2038–2061.

[25] J.B. Martin (2001). Point processes in fast Jackson networks. *The Annals of Applied Probability* **11**, 650-663.

[26] J.B. Martin and Y.M Suhov (1999). Fast Jackson networks. *The Annals of Applied Probability* **9**, 854–870.
[27] M.D. Mitzenmacher (1996). Load balancing and density dependent jump Markov processes. In Proceedings of the Thirty-Seventh Annual Symposium on Foundations of Computer Science, pages 213–222.

[28] M.D. Mitzenmacher (1996). The power of two choices in randomized load balancing. PhD thesis, University of California at Berkeley, Department of Computer Science, Berkeley, CA, 1996.

[29] M. Mitzenmacher (1998). Analyses of load stealing models using differential equations. In Proceedings of the Tenth ACM Symposium on Parallel Algorithms and Architectures, pages 212–221.

[30] M. Mitzenmacher (1999). On the analysis of randomized load balancing schemes. Theory of Computing Systems 32, 361–386.

[31] M. Mitzenmacher (1999). Studying balanced allocations with differential equations. Combinatorics, Probability, and Computing 8, 473–482.

[32] M. Mitzenmacher (2000). How useful is old information? IEEE Transactions on Parallel and Distributed Systems 11, 6–20.

[33] M. Mitzenmacher (2001). The power of two choices in randomized load balancing. IEEE Transactions on Parallel and Distributed Computing 12, 1094-1104.

[34] M. Mitzenmacher, A. Richa, and R. Sitaraman (2001). The power of two random choices: a survey of techniques and results. In Handbook of Randomized Computing: Volume 1, P. Pardalos, S. Rajasekaran and J. Rolim (eds), pages 255-312.

[35] M. Mitzenmacher and B. Vöcking (1998). The asymptotics of selecting the shortest of two, improved. In Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing, pages 326–327.

[36] R. Mirchandaney, D. Towsley, and J.A. Stankovic (1989). Analysis of the effects of delays on load sharing. IEEE Transactions on Computers 38, 1513–1525.

[37] Y.M. Suhov and N.D. Vvedenskaya (2002). Fast Jackson Networks with Dynamic Routing. Problems of Information Transmission 38, 136{153.
[38] B. Vöcking (1999). How asymmetry helps load balancing. In Proceedings of the Fortieth Annual Symposium on Foundations of Computer Science, pages 131–140.

[39] N.D. Vvedenskaya, R.L. Dobrushin and F.I. Karpelevich (1996). Queueing system with selection of the shortest of two queues: An asymptotic approach. Problems of Information Transmissions 32, 20–34.

[40] N.D. Vvedenskaya and Y.M. Suhov (1997). Dobrushin’s mean-field approximation for a queue with dynamic routing. Markov Processes and Related Fields 3, 493–526.

[41] R. Weber (1978). On the optimal assignment of customers to parallel servers. Journal of Applied Probabilities 15, 406–413.

[42] W. Winston (1977). Optimality of the shortest line discipline. Journal of Applied Probabilities 14, 181–189.

[43] N.C. Wormald (1995). Differential equations for random processes and random graphs. The Annals of Applied Probability 5, 1217–1235.

[44] T. Yang and J.G.C. Templeton (1987). A survey on retrial queues. Queueing Systems, 2, 201–233.