Approximating the Performance of a Last Mile Transportation System

By

Hai Wang

Bachelor of Science, Tsinghua University, 2009

Submitted to Sloan School of Management, and Department of Civil and Environmental Engineering, in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Operations Research and

Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

Signature of Author...

Operations Research Center

Department of Civil & Environmental Engineering

May 16th, 2012

Certified by ....

Amedeo R. Odoni

Professor of Aeronautics and Astronautics & Civil and Environmental Engineering

Thesis Supervisor

Accepted by ....

Patrick Jaillet

Co-director, Operations Research Center

Accepted by ....

Heidi M. Nepf

Chair, Departmental Committee for Graduate Students

Approximating the Performance of a Last Mile Transportation System

By

Hai Wang

Submitted to Sloan School of Management, and Department of Civil and Environmental Engineering,
on May 16th, 2012 in Partial Fulfillment of the Requirements for the Degrees of Master of Science in Operations Research and Master of Science in Transportation

Abstract

The Last Mile Problem (LMP) refers to the provision of travel service from the nearest public transportation node to a home or office. We study the supply side of this problem in a stochastic setting, with batch demands resulting from the arrival of groups of passengers at rail stations or bus stops who request last-mile service. Closed-form bounds and approximations are derived for the performance of Last Mile Transportations Systems as a function of the fundamental design parameters of such systems. An initial set of results is obtained for the case in which a fleet of vehicles of unit capacity provides the Last Mile service and each delivery route consists of a simple round-trip between the rail station and bus stop and the single passenger's destination. These results are then extended to the general case in which the capacity of a vehicle is an arbitrary, but typically small (under 10) number. It is shown through comparisons with simulation results, that a particular strict upper bound and an approximate upper bound, both derived under similar assumptions, perform consistently and remarkably well for the entire spectrum of input values and conditions simulated. These expressions can therefore be used for the preliminary planning and design of Last Mile Transportation Systems, especially for determining approximately resource requirements, such as the number of vehicles/servers needed to achieve some pre-specified level of service.

Thesis Supervisor: Amedeo R. Odoni

Title: Professor of Aeronautics and Astronautics & Civil and Environmental Engineering

Acknowledgements

First and foremost, I would like to express my great and sincere appreciation to my advisor Professor Amedeo Odoni. For the last three years, you are the constant source of knowledge, guidance and encouragement for me; you provide me continuous support and endless help all the time. I am significantly benefited from your immerse knowledge, kindly help, inspiration and patience. You are the greatest Professor and mentor I have ever met or known, and I always feel I am so lucky to be your student and work with you. Throughout the research, you provide me sound advice, lots of new ideas, tremendous guidance, and you even helped me with my languages and writing. I could not imagine a better advisor and mentor than you. I really appreciate it.

Also, I would like to thank my parents, my father Zhengqing Wang and my mother Suqing Xu, although you cannot read this acknowledgement in English. You provide the endless support and love all the time. Then years ago, who could imagine your son study and do research in Massachusetts Institute of Technology? I cannot make it without you.

Contents

List of Figures

List of Tables

1. Problem Introduction, Description, Assumptions and Overall Approach

1.1 Introduction and Literature Survey

The Last Mile Problem (LMP) refers to the provision of travel service from home or workplace to the nearest public transportation node ("first mile") or vice versa ("last mile"). This public transportation node could be the nearest rapid transit rail station or a stop of a scheduled bus line. The unavailability of this type of service is one of the main deterrents to the use of public transport in urban areas, especially for certain demographic groups, such as schoolchildren, seniors and the disabled. Currently, the default solutions to the LMP are walking, taking a taxi, or driving a private vehicle.

A conceptual Last Mile Transportation System (LMTS) is described schematically in Figure 1, which shows an urban area surrounding a public-transit rail station, where trains arrive and discharge passengers. The passengers' final destinations (homes, offices and workplaces) are distributed in the area. A fleet of vehicles transports these passengers to their eventual destinations and empty vehicles return to the station to pick up waiting passengers or newly arriving ones. We describe the setting in more detail latter in Chapter 1.2.

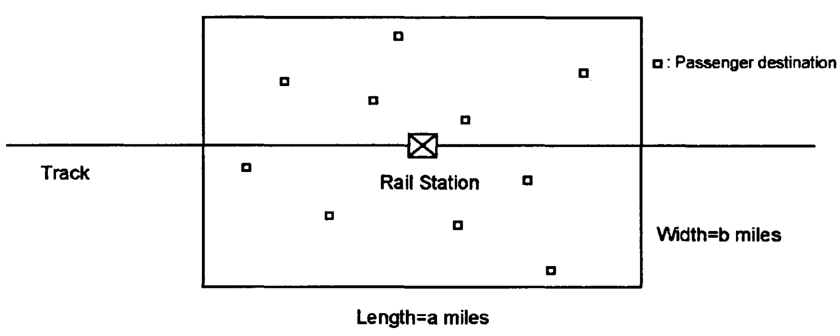


Figure 1: Schematic of a Last Mile Transportation System (LMTS)

Many issues must be addressed when designing and operating a LMTS. On the supply side, it is essential to deal with difficult questions concerning the stochastic aspects of the system. The demand side requires an understanding and estimation of the potential LMTS loads as a function of demographic characteristics, nature of trip, level of service, cost, etc.

The focus of this thesis is solely on the supply side: given a probabilistic description of demand, design a LMTS that operates under dynamic and stochastic conditions according to certain guidelines and satisfies a set of Level of Service (LOS) requirements. This implies specifying such system characteristics as vehicle fleet size, service frequency, dynamically varying vehicle schedules, vehicle dispatching strategies, vehicle routing strategies, monitoring and control of operations, etc.

Addressing these questions is difficult analytically, as the planning and management of a LMTS generally involves such complications as: stochastic travel times that may also change dynamically by time-of-day, according to traffic and weather conditions; batch arrivals of prospective passengers; partitioning of demands among vehicles; routing of the vehicles; queuing issues; and, obviously, numerous considerations concerning staffing and economic sustainability. With the exception of staffing and economic issues, we address most of these complications in this thesis in a static setting.

An extensive literature in this general area has generated various models for a number of application contexts related to the LMP with early papers dating back to the 1970s. We mention here only a few that are among the most influential in the field, as well as relevant to the approach we have adopted.

The Dynamic Traveling Repairman Problem (DTRP) was introduced in two papers by Bertsimas and Van Ryzin. They consider the DTRP in the cases of a single-vehicle "fleet" [1] and of multiple vehicles [2]. The Dynamic Pick-up and Delivery Problem (DPDP) was studied by Swihart and Papastavrou [3], who derived bounds on the performance of several DPDP variants for light and heavy traffic. The Car Pooling Problem (CPP), introduced by Baldacci, Maniezzo and Mingozzi [4] also has features similar to the LMP – or, more exactly, to the First Mile Problem. This paper presents both exact and heuristic methods for solving the CPP based on integer programming formulations. Finally, a large number of papers have dealt with the Dial-a-Ride Problem (DARP) – see, e.g., Jaw, Odoni, Psaraftis and Wilson [15]. A fine critical review of the DARP literature by Cordeau and Laporte [5] underlines, among other points, the fact that this body of work does not address well some of the queuing aspects of the subject systems – a deficiency that this thesis tries to remedy.

It should also be noted that similarities exist between the LMP and various queuing, dispatching, routing, and resource allocation problems arising in entirely different contexts such as the design of manufacturing systems, the operation of elevator banks, and the scheduling of school-bus systems.

The major difference between the LMP and the more "traditional" problems identified above is that, in the LMP, passengers arrive in (possibly large) batches, not singly. Moreover, the size of these batches is a random variable. Queuing systems with batch arrivals are notoriously difficult analytically. A further complication is that the "service times" of passengers are determined by the length (or the duration) of the routes traveled by the fleet of delivery vehicles. Thus, in designing a LMTS, it is necessary to consider simultaneously the problems of: allocating passengers among vehicles; routing the vehicles and estimating the lengths of the routes; and computing the queuing performance characteristics of the system.

The main body of this thsis is organized as follows. In latter Chapter 1, we describe in more detail the version of the LMP problem that we are studying and discuss the associated fundamental assumptions. It will be seen that the problem analyzed is quite generic and that by relaxing one or more of the assumptions, one can capture a broad range of interesting variations. Then we outlines the overall approach utilized to derive our results: we begin by deriving a set of queuing results by considering a fleet of vehicles with capacity for a single passenger ($c = 1$) and then extend the analysis by allowing the vehicle capacity to be arbitrary and by incorporating the resulting travel time estimates into the queuing expressions derived for the $c = 1$ case. Chapter 2 presents our analysis and results for the single-capacity case. We derive three different approximate expressions for queuing performance as a function of the design parameters of the LMTS and then identify, through a set of simulation experiments, the expression that performs best – and, in fact, approximates very well the observed waiting times. Chapter 3 first derives approximate analytical expressions for the travel times associated with fleets consisting of vehicles with a capacity of up to 20 passengers and then applies the queuing approximation derived in Chapter 2 to the multi-passenger capacity case. The results again compare well with those obtained from a simulation. The main part of Chapter 2 and 3 contain only outlines of the lengthy derivations of our results. A sequence of technical sections provides the details following the corresponding main part. Finally, Chapter 4 contains a summary and concluding remarks.

1.2 Problem Description and Assumptions

We now describe in more detail the LMP scenario of Figure 1. The Last Mile Transportation System (LMTS) would operate as follows: Let STA be the transit rail station served by the LMTS and consider a passenger, PAX, who will board a train at station ORIGIN for the purpose of traveling to STA and will then board a LMTS vehicle for transport to her home. PAX will be required to provide advance notice to LMTS of her impending arrival at STA. The time interval between the advance notice and the actual arrival of PAX at STA will be of the order of several minutes (e.g., at least 5 or 10 minutes) to give the LMTS system sufficient time to plan the service of PAX. In practical terms, the advance notice could be generated by PAX in a number of alternative ways. For example, PAX could use a smart-phone when she arrives at ORIGIN or when she enters her train to STA; or, she could tap a smart card on a special-purpose screen, as she is entering ORIGIN or while aboard the train. The resulting message to the LMTS will include the expected time of arrival of PAX at STA (easy to predict, once the passenger is at the ORIGIN station or aboard a train) and her ultimate destination, e.g., her home address. (If the great majority of LMTS users will be subscribers whose home addresses will be pre-registered on a file, then the only information that PAX would have to provide will be an identification number.)

Once the information about PAX is received the LMTS will assign PAX to one of the vehicles of the LMTS fleet, plan the route of that vehicle so it includes a visit to the ultimate destination of PAX, estimate the departure time of the vehicle from STA, and notify PAX accordingly. PAX will receive a message (on her smart-phone or by tapping her card on a screen when she arrives at STA) that indicates the vehicle she has been assigned to and the planned departure time of the vehicle from STA (e.g., "please board Vehicle 123 which will depart from STA at 4:26 PM"). Once all the passengers assigned to a vehicle are on board, the vehicle will execute a delivery route, visiting the destination of each of the passengers and will then return to STA to pick up the passengers for its next delivery tour.

The LMTS described above may be difficult to implement due to many practical issues and considerations. However, we have chosen to study it because it possesses the generic system features that we are most interested in: arrivals of passengers in "batches" (groups) at STA; "real-time" clustering of passengers for assignment to a fleet of vehicles; "real-time" routing of the vehicles to deliver the passengers on board; and fast computation of waiting times and other performance parameters so that, for example, passengers can be notified in a timely way of the departure time of the vehicle they have been assigned to/ informed of the expected departure times and intended use of the LMTS. Actual implementations would involve some simpler variants of the above features.

Given the service region geometry, passenger demand rates, the spatial distribution of the passenger destinations, and the number, capacity and travel speed of the LMTS vehicles, examples of performance metrics that we eventually wish to compute include: the average waiting time until boarding a delivery vehicle, the average riding time of passengers, the average waiting time until delivery, the minimum number of vehicles we need to reach stable operation, vehicle productivity and workload, and eventually (but not in this thesis) the general cost of operating the system and various service vs. cost trade-offs.

We now identify briefly the specifics of the model considered. With reference to Figure 1, we make the following assumptions: (i) headways, *h*, between arrivals of trains at the station (and discharges of passengers) are constant; (ii) passengers are discharged in batches after each train's arrival; (iii) the batch size is a general random variable, ξ, with known expected value, $E(\xi) = \lambda$, and variance, $Var(\xi) = \sigma_\xi^2$; (iv) all passengers arriving in a single batch request service essentially simultaneously; (v) given the size of any particular batch, $\xi = \xi_0$, the destinations of the ξ_0 passengers in the batch are distributed identically, uniformly and independently in a service region; (vi) the service region is convex and compact with known dimensions; (vii) the delivery fleet (or pick-up fleet, in the case of "First Mile" service) consists of m vehicles, each with integer capacity, c.

We believe that (i) – (vii) are adequately general assumptions for approximating, to a first order, the characteristics of many potential variations of LMTS. Note that our model includes the most difficult, from the analytical point of view, features that one might encounter in an LMTS: batch arrivals, stochastic demand, stochastic service times, and the presence of queuing phenomena interfaced with routing problems.

To ensure that the mathematical expressions presented in Chapter 2 and 3 below are adequately concise, we have also used the following three simplifying assumptions: (viii) the service area, where the destinations of the passengers are located, is a $b \times b$ square, with the train station, STA, located at the square's center; (ix) the travel medium is continuous, homogeneous, and planar; and (x) the travel speed is constant throughout the service region and equal to 1. We have studied a number of variants of assumptions (viii) and (ix), such as cases in which the region is not a square, or the travel metric is Euclidean or rectangular ("right-angle) or contains discontinuities (e.g., barriers to travel), and shown that such mild changes in the assumptions pose no particular challenges.

1.3 Description of Overall Approach

Chapter 2 and 3 of the thesis describe in detail our analysis and results. In this Chapter we provide a brief description of the overall approach we have followed to provide perspective for these detailed Chapters. We have adopted a perspective under which the LMTS is regarded as a spatially distributed queuing system in which the demands are as described before (batch arrivals of passengers with a constant headway between the arrivals of successive batches). In line, with typical queuing terminology, we shall refer henceforth to passengers as "*customers*" of the spatially distributed queuing system. The m parallel servers (the vehicle fleet) serve customers in groups of c or smaller, where c is the capacity of each vehicle. The service time for each group is equal to the travel time associated with a vehicle tour that begins at the station/depot, visits each of the c (or fewer) customer destinations and returns to the station/depot to pick up a new group.

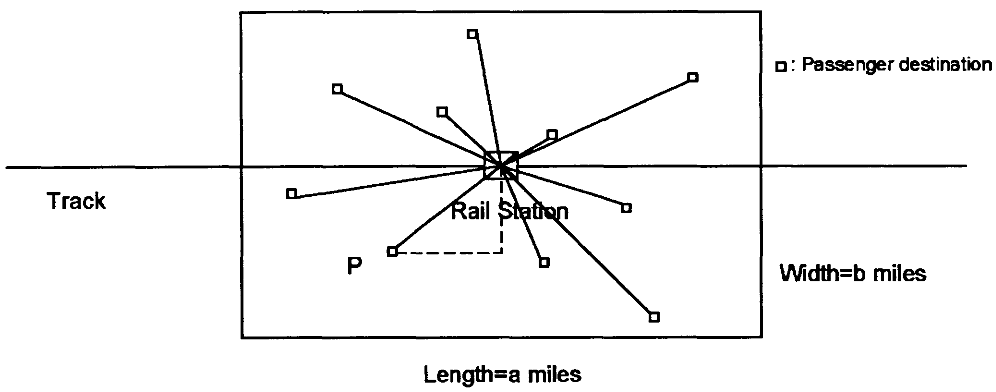


Figure 2: Customer destinations and vehicles routes of the Unit-Capacity, Multi-Vehicle LMP

Because queuing systems with batch arrivals (like the arrivals of passengers at STA) are notoriously difficult to analyze, we resort to a two-step approach. In Step 1, we assume that $c = 1$, i.e., that the delivery vehicles have unit capacity. Thus, in this case, service times consist simply of the duration of a round-trip between STA and one passenger's destination (see Figure 2), with the destination being randomly and uniformly distributed within the service area per our assumption (v) in Chapter 1. In this way we obtain a $D^{\xi}/G/m/\infty$ system in queuing theory notation, where: D^{ξ} indicates batch arrivals at constant ("Deterministic") intervals with the number of arriving passengers in each batch described by random variable ξ; G denotes the fact that the distribution of service times (i.e., the duration of the round trips between STA and customer destinations) is "general"; and m and ∞ indicate, respectively, the number of service vehicles and the fact that no *a priori* limit is placed on the number of customers waiting for pickup at STA.

As no closed-form expressions are available for the fundamental quantities the performance of a $D^{\xi}/G/m/\infty$ system, we then attempt to obtain expressions that would help us estimate performance by studying similar queuing systems, which are simpler to analyze mathematically. In this way, and through a series of simplifications, we derive one lower bound and two upper bounds for the mean waiting time associated with $D^{\xi}/G/m/\infty$ queues. We then carry out an extensive series of simple simulation experiments and conclude that one of these three approximations (an upper bound) provides very good estimates of the performance of the system under a broad range of system design parameters. We therefore adopt this approximate expression for studying the general vehicle capacity case in which c can take on any (usually small) integer value.

Step 2 examines this general case, in which service times are equal to the duration of delivery tours consisting of $c(> 1)$ or fewer delivery stops, as shown in Figure 3. To

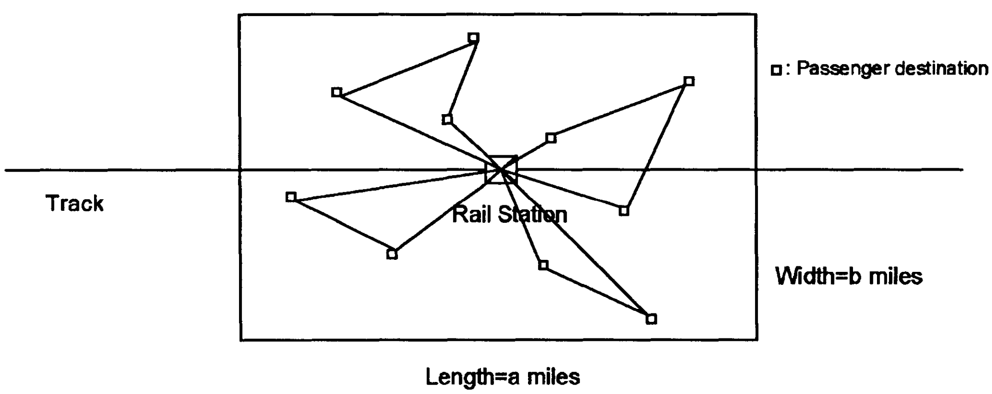


Figure 3: Vehicle routes of the General-Capacity, Multi-Vehicle LMP

apply to the general capacity case the queuing expressions that were derived in Step 1, we need to compute in Step 2, the approximate length and the variance of the length of the vehicle tours shown in Figure 3. We accomplish this by using arguments from geometrical probability and from the literature on the Traveling Salesman Problem. We obtain several such approximate expressions in this way and compare them with the results of another series of simple simulation experiments to select the expressions that fit best the observed expected values and variances of the vehicle tour lengths. We then use these expressions, along with the queuing-based approximation derived in Step 1, to complete the process of estimating the performance of the LMTS for the general case of arbitrary fleet size and arbitrary vehicle capacity.

The main part of Chapter 2 and 3 provide only an outline of the (occasionally lengthy) derivations of the results contained therein. The detailed mathematics is after the outline and between dash line.

2. The Unit-Capacity, Multi-Vehicle LMP

In this Chapter we consider the analysis of the Unit-Capacity, Multi-Vehicle case, described in Chapter 1 as Step 1, in which $c = 1$, and m is an arbitrary positive integer. As already indicated above (Figure 2), the length of the vehicle trips in this case is equal to two times the distance between the rail station and a customer's destination. For the purpose of keeping relatively simple the various expressions derived, and without loss of generality, we shall assume that travel in the rectangular region of interest [Assumption (viii) in Chapter 1] is according to the right-angle metric, with directions of travel parallel to the sides of the rectangle. A typical route, for serving a particular customer P is indicated through a dashed line in Figure 2. Because we have also postulated [Assumption (x)] constant and unit travel speeds, the expressions for travel times in the region are identical with those derived for travel distances.

The basic notation is summarized as follows:

h = the constant headway between arrivals of trains at the station STA (and discharges of customers);

ξ = a random variable denoting the number of LMTS customers ("batch size") discharged after the arrival of a train at STA – with the sizes of successive batches being mutually independent and with $E(\xi) = \lambda$, and $Var(\xi) = \sigma_\xi^2$ denoting, respectively, the expected value and variance of ξ;

S = a random variable denoting the service time of any random LMTS customer with $E(S) = s$ and variance $Var(S)$;

Note that the successive service times by any given vehicle in the fleet are independent and identically distributed. The traffic load (or utilization ratio) is given by $\rho = s\lambda/mh$. Note that m/s is the service rate of the LMTS, while λ/h is the rate of customer arrivals per unit of time. Technical Section 2.T presents some background results that are useful in the analysis of the Unit-Capacity, Multi-Vehicle case.

2.T*: Background Results

1. Expectation and variance of composite random variables;

Given a sequence of independent random variables $X_i(i = 1,2,\ldots,N)$, where N is also a random variable and independent of all the X_i, let $Y = \sum_{i=1}^{N} X_i$.

It is known [6] that:

$$E(Y) = E(N)E(X) \quad (2.T.1)$$

$$Var(Y) = Var(N)E^2(X) + E(N)Var(X) \quad (2.T.2)$$

where $E(Y), E(N)$ and $E(X)$ denote the expected values, and $Var(Y), Var(N)$ and $Var(X)$ denote the variances of Y, N and X, respectively.

2. Total expectation and total variance;

Given two random variables X and Y, it is known [6] that:

$$E(X) = E(E(X|Y)) \quad (2.T.3)$$

$$Var(X) = E\big(Var(X|Y)\big) + Var(E(X|Y)) \quad (2.T.4)$$

3. Exact solution for average waiting time in a M/G/1/∞ queue;

In a $M/G/1/\infty$ queue, using the same notations as in the main part, it is known [7] that:

$$W_{M/G/1/\infty} = \frac{\rho}{1-\rho} \cdot \frac{1 + C_s^2}{2} s \quad (2.T.5)$$

4. Bound for the average waiting time in a GI/G/1/∞ queue;

In a $GI/G/1/\infty$ queue, let $1/\lambda_a$ and s be the expected inter-arrival time and service time, respectively, $\mu = 1/s$ the service rate, σ_a^2, σ_s^2 the variances of the inter-arrival time and service time, respectively, $C_a^2 = \lambda_a^2\sigma_a^2, C_s^2 = \sigma_s^2/s^2$ the coefficients of variation of the inter-arrival time and service time, respectively, and $\rho = \lambda_a s = \lambda_a/\mu$ the traffic load (system utilization ratio). There is no simple explicit expression for the average waiting time W. According to [7],

$$\frac{\rho^2(1+C_s^2)-2\rho}{2\lambda_a(1-\rho)} \le W_{GI/G/1/\infty} \le \frac{\lambda_a(C_a^2+C_s^2)}{2(1-\rho)} \tag{2.T.6}$$

The upper bound becomes asymptotically exact as $\rho \longrightarrow 1$.

5. Approximation for average waiting time in a GI/G/1/∞ queue:

In a $GI/G/1/\infty$ queue, using a combination of queuing theoretic and numerical analysis, the following two-moment approximation for the average waiting time in queue per customer was obtained by Kramer and Langenbach-Belz [8]:

$$W_{GI/G/1/\infty} \approx \frac{\rho}{1-\rho}\cdot\frac{(C_a^2+C_s^2)J}{2}s, \text{where } J = \begin{cases} \exp\left[-\frac{2(1-\rho)(1-C_a^2)^2}{3\rho(C_a^2+C_s^2)}\right], C_a^2 \le 1; \\ \exp\left[-\frac{(1-\rho)(C_a^2-1)}{C_a^2+4C_s^2}\right], C_a^2 > 1; \end{cases} \tag{2.T.7}$$

The approximation is useful for practical purposes provided that the traffic load of the system is not small and C_a^2 is not too large. In the LMP, we will choose the number of vehicles so as to make sure the system utilization ratio (traffic load) is not small. Additionally, the constant headway of successive batch arrival (see the problem definition in the main part) means that $C_a^2 = 0$. Therefore, the approximation could work well for the LMP.

2.1 A Lower Bound

We are particularly interested in the expected waiting time, W, of LMTS customers until they board one of the m vehicles to be transported to their eventual destination. Determining this expected waiting time, as a function of the LMTS design parameters is a critical step toward developing the means to design LMTS satisfying certain level-of-service requirements. We begin by obtaining a lower bound for W.

Since no exact analytical solution exists for the complicated $D^{\xi}/G/m/\infty$ queuing model, we consider a modified system in which, instead of having batch arrivals with average size $E(\xi)$ at constant intervals (headway = h), we have a single arrival of a customer every $h/E(\xi)$ units of time. This modification transforms the original $D^{\xi}/G/m/\infty$ system into a $D/G/m/\infty$ queuing system. The latter is characterized by a shorter average waiting time, W, than the original $D^{\xi}/G/m/\infty$ system since the arrivals of customers are deterministic and evenly distributed, while the total expected number of customers served by the two systems is the same. However, no exact analytical solution exists for the $D/G/m/\infty$ model either. Therefore, we consider instead a $D/G/1/\infty$ model, which has identical customer inter-arrival times with the $D/G/m/\infty$ model, while its single server works m times faster than each of the servers of the m-server system. Following the "remaining work inequality" principle of multi-server queuing models in [9] and applying the approximation of $GI/G/1/\infty$ given in [7] (see Technical section 2.2.T) we can then obtain (Technical Section 2.1.T1) a lower bound as follows:

$$W \geq \frac{E(\xi)E(S)E(S^2) + hE(S^2) - 2hE^2(S) - mhE(S^2)}{2E(S)\big(mh - E(\xi)E(S)\big)} \tag{1}$$

when the size of customer arrival batches, ξ, is drawn from a General distribution and the customer service time, S, is also drawn from a General distribution.

For the special case (Technical Section 2.1.T2 and 2.1.T3) in which the size of customer arrival batches is a Poisson random variable with intensity λ and the service region is a $b \times b$ square:

$$W \geq \frac{-7mbh + 7bh + 7b^2\lambda}{12(mh - b\lambda)} \tag{2}$$

2.1.T1*: Lower Bound of Unit-Capacity, Multi-Vehicle LMP in the general case

According to the "remaining work inequality" for multi-server queuing model in [9], for the $D/G/m/\infty$ model and the corresponding $D/G/1/\infty$ model, constructed in the way described in main part Chapter 2.1, we have the following inequality:

$$W_{D/G/m/\infty} \geq W_{D/G/1/\infty} - \frac{\mu(m-1)(\sigma_s^2 + 1/\mu^2)}{2m} \tag{2.1.T1.1}$$

For the $D/G/1/\infty$ model, the average service time is reduced to $s = E(S)/m$, the service rate $\mu = m/E(S)$, the service time variance $\sigma_s^2 = Var(S)/m^2$, the coefficient of variation $C_s^2 = (Var(S)/m^2)/(E^2(S)/m^2) = Var(S)/E^2(S)$, and the queue utilization ratio $\rho = E(\xi)E(S)/mh$.

According to [7]:

$$W_{D/G/1/\infty} \geq \frac{\rho^2(1+C_s^2)-2\rho}{2\lambda_a(1-\rho)} = \frac{\left(\frac{E(\xi)E(S)}{mh}\right)^2\left(1+\frac{Var(S)}{E^2(S)}\right)-2\frac{E(\xi)E(S)}{mh}}{2\frac{E(\xi)}{h}(1-\frac{E(\xi)E(S)}{mh})}$$
$$= \frac{E^2(\xi)E^2(S)+Var(S)E^2(\xi)-2mhE(\xi)E(S)}{2mE(\xi)(mh-E(\xi)E(S))} \quad (2.1.T1.2)$$

For the $D/G/m/\infty$ model, the service rate $\mu = 1/E(S)$ and the service time variance $\sigma_s^2 = Var(S)$. We then have:

$$W_{D/G/m/\infty} \geq \frac{E^2(\xi)E^2(S)+Var(S)E^2(\xi)-2mhE(\xi)E(S)}{2mE(\xi)(mh-E(\xi)E(S))} - \frac{\frac{1}{E(S)}(m-1)\left(Var(S)+E^2(S)\right)}{2m}$$
$$= \frac{E^2(\xi)E^2(S)+Var(S)E^2(\xi)-2mhE(\xi)E(S)}{2mE(\xi)(mh-E(\xi)E(S))} - \frac{(m-1)E(S^2)}{2mE(S)}$$
$$= \frac{E(\xi)E(S)E(S^2)+hE(S^2)-2hE^2(S)-mhE(S^2)}{2E(S)(mh-E(\xi)E(S))} \quad (2.1.T1.3)$$

This is the strict lower bound for the average waiting time in the original $D^{\xi}/G/m/\infty$ model.

Under heavy traffic [9],

$$W_{D/G/1/\infty} = \frac{\lambda_a(\sigma_s^2+\sigma_a^2)}{2(1-\rho)} = \frac{\frac{E(\xi)}{h}\frac{Var(S)}{m^2}}{2(1-\frac{E(\xi)E(S)}{mh})} = \frac{Var(S)E(\xi)}{2m(mh-E(\xi)E(S))}$$
$$= \frac{E(S^2)E(\xi)-E^2(S)E(\xi)}{2m(mh-E(\xi)E(S))} \quad (2.1.T1.4)$$

$$W_{D/G/m/\infty} \geq \frac{E(S^2)E(\xi)-E^2(S)E(\xi)}{2m(mh-E(\xi)E(S))} - \frac{(m-1)E(S^2)}{2mE(S)} =$$
$$= \frac{mE(\xi)E(S)E(S^2)+mhE(S^2)-m^2hE(S^2)-E^3(S)E(\xi)}{2mE(S)(mh-E(\xi)E(S))} \quad (2.1.T1.5)$$

This is the strict lower bound for the average waiting time under heavy traffic in the original $D^{\xi}/G/m/\infty$ model.

2.1.T2*: Service time distribution in a rectangular service region

Assume a rectangular service region Λ with dimensions of a along the horizontal axis and b along the vertical axis and with $a \geq b$. We also assume a right-angle ("Manhattan") travel metric with the directions of travel parallel to the sides of the rectangle. The train station is located at the center of the rectangular area and it is also the origin of our system coordinates. The maximum travel distance required to deliver a customer to his/her final destination and return to the station is $a + b$, while the minimum travel distance is 0.

Since the customer destinations are uniformly and independently distributed within the area and vehicles travel with unit velocity, successive travel times along the X-axis are uniformly and independently distributed in $[0, a]$ with probability density function $f_X(x) = 1/a, 0 \leq x \leq a$; similarly, travel times along the Y-axis are uniformly and independently distributed in $[0, b]$ with probability density function $f_Y(y) = 1/b, 0 \leq y \leq b$.

Therefore, the total travel time $S = X + Y$, is described by the following probability density function:

$$f_S(s) = \begin{cases} \dfrac{1}{ab}s, & 0 \leq s \leq b; \\ \dfrac{1}{a}, & b \leq s \leq a; \\ \dfrac{1}{a} + \dfrac{1}{b} - \dfrac{1}{ab}s, & a \leq s \leq a + b; \end{cases}$$

with

$$E(S) = \frac{a+b}{2}, Var(S) = \frac{a^2 + b^2}{12}, E(S^2) = \frac{2a^2 + 2b^2 + 3ab}{6}$$

When the region is a square, i.e., $a = b$,

$$E(S) = b, Var(S) = \frac{b^2}{6}, E(S^2) = \frac{7b^2}{6}$$

In the analysis above, we ignored any time required for loading and unloading customers.

2.1.T3*: Lower Bound of the Unit-Capacity, Multi-Vehicle LMP for a Poison customer batch size and a square service region

If the number of customers from each train is Poisson-distributed and the service time is as the square service region case described in Technical Section 2.1.T2, we consider a modified system in which, instead of having batch Poisson arrivals at the rate of λ at constant intervals (headway $=h$), we have a continuous Poisson arrival stream at the rate of $\lambda_l = \lambda/h$ per unit of time. Both the original and modified systems have the same overall average arrival rate of λ customers every h time units.

Considering the corresponding single (but m times faster) server model, the average service time is reduced to $s = E(S)/m = b/m$, the service rate $\mu = m/b$, the service time variance $\sigma_s^2 = Var(S)/m^2 = b^2/6m^2$, the coefficient of variation of the service time $C_s^2 = \sigma_s^2/s^2 = 1/6$, and the queue utilization ratio $\rho = \lambda_l/(m/b) = b\lambda/mh$. Thus:

$$W_{M/G/1/\infty} \geq \frac{\rho}{1-\rho} \cdot \frac{1+C_s^2}{2} s = \frac{\frac{b\lambda}{mh}}{1-\frac{b\lambda}{mh}} \cdot \frac{1+1/6}{2} \cdot \frac{b}{m} = \frac{7}{12} \cdot \frac{\frac{b\lambda}{mh}}{1-\frac{b\lambda}{mh}} \cdot \frac{b}{m} = \frac{7b^2\lambda}{12m(mh-b\lambda)}$$

For the $M/G/m/\infty$ model, the service rate $\mu = 1/b$, and the service time variance $\sigma_s^2 = Var(S)/m^2 = b^2/6$. Thus:

$$\begin{aligned}
W_{M/G/m/\infty} &\geq \frac{7b^2\lambda}{12m(mh-b\lambda)} - \frac{\mu(m-1)(\sigma_s^2 + \frac{1}{\mu^2})}{2m} = \frac{7b^2\lambda}{12m(mh-b\lambda)} - \frac{(m-1)(\frac{b^2}{6} + b^2)}{2mb} \\
&= \frac{7b^2\lambda}{12m(mh-b\lambda)} - \frac{7(m-1)b}{12m} = \frac{7b^2\lambda - 7(m-1)b(mh-b\lambda)}{12m(mh-b\lambda)} \\
&= \frac{7b^2\lambda - 7m^2bh + 7mbh + 7mb^2\lambda - 7b^2\lambda}{12m(mh-b\lambda)} = \frac{-7mbh + 7bh + 7b^2\lambda}{12(mh-b\lambda)}
\end{aligned}$$

This is the strict lower bound for the average waiting time in the original $D^{\xi}/G/m/\infty$ model.

Note that the expression above is correct dimensionally, with the dimension (unit) of the expression is first power of time.

2.2 Two Upper Bounds

We next turn to obtaining an upper bound for W in the original Unit-Capacity, Multi-Vehicle $D^{\xi}/G/m/\infty$ model. To do this, we pre-assign customers to different vehicles and construct a corresponding single-server queuing model $D^{N}/G/1/\infty$ for each vehicle, where N is the random variable indicating the number of customers from a single train assigned to the same vehicle.

With such an assignment policy, service inefficiencies exist since a customer is required to wait for his or her assigned vehicle, even when other vehicles may be available. Thus, the average waiting time in this case will be larger than the average waiting time in the original model and provides an upper bound. The customer flow is shown schematically in Figure 4 below.

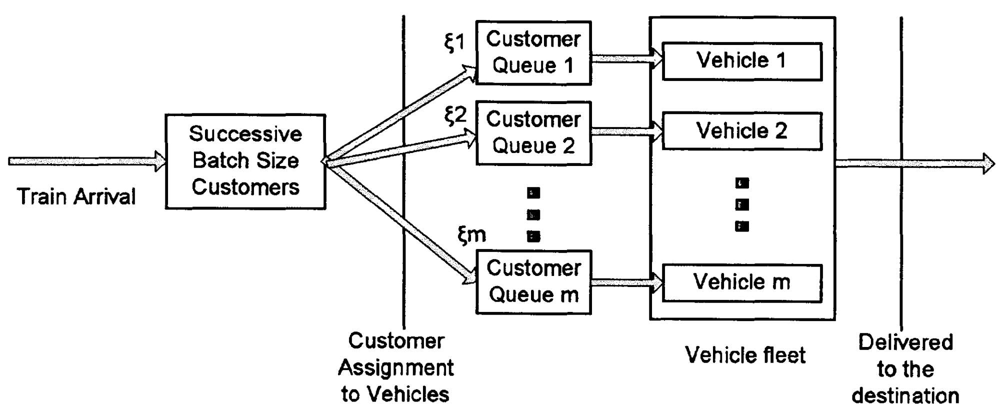


Figure 4: Customer flow in the pre-assignment policy

The $D^N/G/1/\infty$ model is still difficult to work with. To obtain approximate expressions for W, we decompose the problem into two parts (Technical Section 2.2.T). First, the N customers in some batch who are assigned to the same vehicle are treated as a single "macro-customer" P. This reduces the $D^N/G/1/\infty$ model to the more tractable $D/G/1/\infty$ model and allows us to obtain an upper bound for W_1, the expected waiting time until the first customer in P receives service. In a second step, we then compute the additional expected waiting time, W_2, that the i-th customer in P suffers due to being preceded for service by i-1 other customers in P. Thus, the expected waiting time of a customer P is given by $W = W_1 + W_2$. In Technical Section 2.2.T we show that:

$$W_1 \leq \frac{E(N)Var(S) + E^2(S)Var(N)}{2(h - E(N)E(S))}$$

$$W_2 = \frac{E(S)Var(N) + E(S)E^2(N) - E(S)E(N)}{2E(N)}$$

Thus the upper bound we seek is:

$$W \leq \frac{E(N)Var(S) + E^2(S)Var(N)}{2(h - E(N)E(S))} + \frac{E(S)Var(N) + E(S)E^2(N) - E(S)E(N)}{2E(N)} \tag{3}$$

The bound (3) is valid under general assumptions about the probability density functions of the batch size, ξ, and the service times, S. Moreover, (3) has been derived without considering how exactly customers are assigned to vehicles. We analyze next two different policies for customer assignment to vehicles. Each of these policies will provide different modified $D^N/G/1/\infty$ models with different $E(N)$ and $Var(N)$, leading to different expressions for W_1 and W_2, and, ultimately, different upper bounds for W.

2.2.T*: Upper bound for the average waiting time in the $D^N/G/1/\infty$ queue model

We treat all the customers assigned simultaneously (in the same batch) to any given vehicle as a single "macro customer". If we only consider the macro customer, the $D^N/G/1/\infty$ model can be reduced to a $D/G/1/\infty$ model. We denote the average waiting time of a macro customer in the $D/G/1/\infty$ model by W_M. W_M is exactly equal to the average waiting time of the first customer composited in the macro customer, which is denoted by W_1, i.e., $W_M = W_1$. The service time of the macro customer is $T = \sum_{i=1}^{N} S_i$, where N is positive integer random variable indicating the number of real customers composing the macro customer. N depends on the assignment policy. $S_1, S_2, \ldots, S_N$ are the service times of the real customers and they are mutually independent and identically distributed. Therefore,

$$E(T) = \sum_{i=1}^{N} E(S_i) = E(N)E(S)$$

$$Var(T) = E(N)Var(S) + E^2(S)Var(N)$$

$$C_t^2 = \frac{Var(T)}{E^2(T)} = \frac{E(N)Var(S) + E^2(S)Var(N)}{E^2(N)E^2(S)}$$

$\sigma_a^2 = 0$ (Due to constant batch or macro customer inter-arrival time)

$$\lambda_a = \frac{1}{h}$$

$$\rho = \frac{E(T)}{h} = \frac{E(N)E(S)}{h}$$

According to [9], the upper bound of W_1, the average waiting time until the macro customer receive service, is:

$$W_1 = W_{D/G/1/\infty} \le \frac{\lambda_a(\sigma_a^2 + \sigma_s^2)}{2(1-\rho)} = \frac{\frac{1}{h}(0 + Var(T))}{2(1 - \frac{E(T)}{h})} = \frac{E(N)Var(S) + E^2(S)Var(N)}{2(h - E(N)E(S))} \quad (2.2.T.1)$$

According to [8], an approximation of W_1 is given by:

$$W_1 \approx \frac{\frac{E(T)}{h}}{1 - \frac{E(T)}{h}} \cdot \frac{\frac{Var(T)}{E^2(T)}}{2} \cdot E(T) \cdot \exp\left[-\frac{2\left(1 - \frac{E(T)}{h}\right)}{3\frac{E(T)}{h}\frac{Var(T)}{E^2(T)}}\right]$$

$$= \frac{Var(T)}{2(h - E(T))} \cdot \exp\left[-\frac{2(h - E(T))E(T)}{3Var(T)}\right] \quad (2.2.T.2)$$

Assume we have obtained W_1, given n customers composing the macro customer:

When $n = k, k \geq 1.$, the customer in the $i\ th$ position will suffer the average waiting time $W_{i\,th} = W_1 + \sum_{j=1}^{i-1} s_j$, where s_j is the average service time of the $j\ th$ customer served before the $i\ th$ customer. We know the average service time of every customer is $E(S)$, so:

$$W_{i\,th} = W_1 + (i-1)E(S)$$

Let W_k denote the average waiting time of all the k customers,

$$W_k = \frac{\sum_{i=1}^{k} W_{i\,th}}{k} = \frac{\sum_{i=1}^{k}[W_1 + (i-1)E(S)]}{k} = W_1 + \frac{(k-1)E(S)}{2}, k \geq 1$$

When $n = 0$, no customers served and $W_0 = 0$.

Let $W_{D^N/G/1/\infty}$ denote the average waiting time of all customers in the $D^N/G/1/\infty$ model. According to the Law of Total Expectation:

$$W_{D^N/G/1/\infty} = \frac{\sum_{k=0}^{\infty} P(n=k) W_k k}{\sum_{k=0}^{\infty} P(n=k)k} = \frac{\sum_{k=0}^{\infty} P(n=k)[W_M + \frac{(k-1)E(S)}{2}]k}{\sum_{k=0}^{\infty} P(n=k)k}$$

$$= \frac{\sum_{k=0}^{\infty} P(n=k) W_M k}{\sum_{k=0}^{\infty} P(n=k)k} + \frac{\sum_{k=0}^{\infty} P(n=k)\frac{(k-1)E(S)}{2}k}{\sum_{k=0}^{\infty} P(n=k)k}$$

$$= W_1 + \frac{\sum_{k=0}^{\infty} P(n=k)(k-1)k}{\sum_{k=0}^{\infty} P(n=k)k}\frac{E(S)}{2} = W_1 + \frac{E(N^2) - E(N)}{E(N)}\frac{E(S)}{2}$$

$$= W_1 + \frac{E(S)Var(N) + E(S)E^2(N) - E(S)E(N)}{2E(N)} \qquad (2.2.T.3)$$

That is:

$$W_{D^N/G/1/\infty} \leq \frac{E(N)Var(S) + E^2(S)Var(N)}{2(h - E(N)E(S))} + \frac{E(S)Var(N) + E(S)E^2(N) - E(S)E(N)}{2E(N)}$$

The first part

$$W_1 \leq \frac{E(N)Var(S) + E^2(S)Var(N)}{2(h - E(N)E(S))} \qquad (2.2.T.5)$$

is the average waiting time until the first customer assigned to the vehicle in one batch receives service.

The second part

$$W_2 = \frac{E(S)Var(N) + E(S)E^2(N) - E(S)E(N)}{2E(N)} \qquad (2.2.T.6)$$

is the average waiting time due to the service time of customers served before in the same batch.

2.2.1 Randomized Assignment Policy

One possible policy is to assign all the customers randomly (with equal probability $1/m$) and independently to the m different vehicles, with every vehicle serving individually the stream assigned to it. This is illustrated in Figure 5 below:

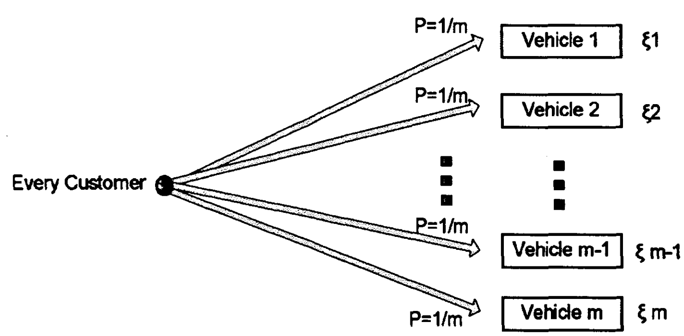


Figure 5: Randomized assignment policy

The model corresponding to the randomized assignment policy led (Technical Section 2.2.1.T1) to the following strict upper bound for the case of a General distribution of customer batch sizes and a general distribution of customer service times:

$$W \leq \frac{mhE(S)E(\xi^2) - mhE(S)E(\xi) + mE(S^2)E^2(\xi) - E^2(S)E^3(\xi)}{2m\big(mh - E(\xi)E(S)\big)E(\xi)} \qquad (4)$$

When the customer batch size is a Poisson random variable and the service region is a $b \times b$ square, the strict upper bound (4) becomes (Technical Section 2.2.1.T1):

$$W \leq \frac{7b^2\lambda m + 6b\lambda mh - 6b^2\lambda^2}{12m(mh - b\lambda)} \qquad (5)$$

An approximate upper bound for the case of Poisson customer batch size and a square service region can also be derived. This last bound was obtained (also in Technical Section 2.2.1.T2) using an approximate expression for the average waiting time of the $GI/G/1/\infty$ queuing model given in [8]:

$$W \le \frac{7b^2\lambda}{12(mh - b\lambda)} \cdot \exp\left[-\frac{4(mh - b\lambda)}{7bm}\right] + \frac{b\lambda}{2m} \tag{6}$$

2.2.1.T1*: Upper bound of the Unit-Capacity, Multi-Vehicle LMP under randomized assignment policy in the general case

One possible assignment policy is to apportion all the customers randomly, uniformly and independently among the m different servers, with each server then serving its own stream of customers independently of all the other servers. The model corresponding to this randomized assignment policy is $D^{N_i}/G/1/\infty$, where N_i is the random variable indicating the number of customers assigned to server i. $N_1, N_2, \ldots, N_m$ are identically distributed, so all $D^{N_i}/G/1/\infty$ models can be taken as the same $D^N/G/1/\infty$ model, although $N_1, N_2, \ldots, N_m$ are not necessarily independent. Assume ξ is the random variable indicating the total number of customers coming from one train. Given ξ, we know $N_1 \sim B(\xi, 1/m)$, in which $B(n,p)$ is Binomial distribution with total number n and individual probability p. Thus:

$$E(N|\xi) = \frac{\xi}{m}$$

$$Var(N|\xi) = \frac{\xi(m-1)}{m^2}$$

$$E(N) = E(E(N|\xi)) = \frac{E(\xi)}{m}$$

$$Var(N) = E(Var(N|\xi)) + Var(E(N|\xi)) = E\left(\frac{\xi(m-1)}{m^2}\right) + Var\left(\frac{\xi}{m}\right)$$
$$= \frac{E(\xi)(m-1)}{m^2} + \frac{Var(\xi)}{m^2} = \frac{E(\xi)(m-1) + Var(\xi)}{m^2}$$

Therefore,

$$W_{D^N/G/1/\infty} \le \frac{E(N)Var(S) + E^2(S)Var(N)}{2(h - E(N)E(S))} + \frac{E(S)Var(N) + E(S)E^2(N) - E(S)E(N)}{2E(N)}$$

$$\le \frac{\frac{E(\xi)}{m}Var(S) + E^2(S)\frac{E(\xi)(m-1) + Var(\xi)}{m^2}}{2\left(h - \frac{E(\xi)}{m}E(S)\right)}$$

$$+ \frac{E(S)\frac{E(\xi)(m-1) + Var(\xi)}{m^2} + E(S)\left(\frac{E(\xi)}{m}\right)^2 - E(S)\frac{E(\xi)}{m}}{2\frac{E(\xi)}{m}}$$

$$\le \frac{mE(\xi)Var(S) + (m-1)E(\xi)E^2(S) + E^2(S)Var(\xi)}{2m(mh - E(\xi)E(S))}$$

$$+ \frac{E(S)(Var(\xi) - E(\xi) + E^2(\xi))}{2mE(\xi)}$$

$$= \frac{mE(\xi)E(S^2) + E^2(S)\left(E(\xi^2) - E^2(\xi) - E(\xi)\right)}{2m(mh - E(\xi)E(S))} + \frac{E(S)\left(E(\xi^2) - E(\xi)\right)}{2mE(\xi)}$$

$$= \frac{mhE(S)E(\xi^2) - mhE(S)E(\xi) + mE(S^2)E^2(\xi) - E^2(S)E^3(\xi)}{2m(mh - E(\xi)E(S))E(\xi)} \quad (2.2.1.T1.1)$$

This is the strict upper bound for the average waiting time under randomized assignment policy in the general case.

2.2.1.T2*: Upper bound of the Unit-Capacity, Multi-Vehicle LMP under randomized assignment policy for a Poisson customer batch size and a square service region

If the number of customers from each train is Poisson-distributed and the service time is as the square service region case described in Technical Section 2.1.T2. If we use the randomized assignment policy, all the customers are randomly and uniformly assigned to m different servers. It is well known that if we assign each customer independently to server j with probability $1/m$ for all j, then the resulting size of each stream will follow identical Poisson distribution with intensity $\lambda_r = \lambda/m$.

$E(\xi) = \lambda$

$$Var(\xi) = \lambda$$

$$E(\xi^2) = Var(\xi) + E^2(\xi) = \lambda + \lambda^2$$

$$E(S) = b$$

$$Var(S) = \frac{b^2}{6}$$

$$E(S^2) = Var(S) + E^2(S) = \frac{7b^2}{6}$$

$$E(N) = \frac{\lambda}{m}$$

$$Var(N) = \frac{\lambda}{m}$$

$C_a^2 = 0$ (Due to constant batch or macro customer inter-arrival time)

$$C_s^2 = C_t^2 = \frac{E(N)Var(S) + E^2(S)Var(N)}{E^2(N)E^2(S)} = \frac{\frac{\lambda}{m}\frac{b^2}{6} + b^2\frac{\lambda}{m}}{\frac{\lambda^2}{m^2}b^2} = \frac{7m}{6\lambda}$$

$$s = E(T) = E(N)E(S) = \frac{b\lambda}{m}$$

$$\rho = \frac{E(T)}{h} = \frac{E(N)E(S)}{h} = \frac{b\lambda}{mh}$$

Thus, using the conclusion of Technical Section 2.2.T, we can obtain a strict upper bound for the average waiting time in the original $D^\xi/G/m/\infty$ model:

$$W \le \frac{mhE(S)E(\xi^2) - mhE(S)E(\xi) + mE(S^2)E^2(\xi) - E^2(S)E^3(\xi)}{2m\big(mh - E(\xi)E(S)\big)E(\xi)}$$

$$= \frac{mhb(\lambda + \lambda^2) - mhb\lambda + m\frac{7b^2}{6}\lambda^2 - b^2\lambda^3}{2m(mh - \lambda b)\lambda} = \frac{7b^2\lambda m + 6b\lambda mh - 6b^2\lambda^2}{12m(mh - b\lambda)}$$

Similarly, we have the approximation (using the same notations as in Technical Section 2.2.T):

$$W_1 \approx \frac{\rho}{1-\rho} \cdot \frac{(C_a^2 + C_s^2)}{2} s \cdot \exp\left[-\frac{2(1-\rho)(1-C_a^2)^2}{3\rho(C_a^2 + C_s^2)}\right]$$

$$= \frac{\frac{b\lambda}{mh}}{1-\frac{b\lambda}{mh}} \cdot \frac{\frac{7m}{6\lambda}}{2}\frac{b\lambda}{m} \cdot \exp\left[-\frac{2\left(1-\frac{b\lambda}{mh}\right)}{3\frac{b\lambda}{mh}\frac{7m}{6\lambda}}\right]$$

$$= \frac{7b^2\lambda}{12(mh-b\lambda)} \cdot \exp\,[-\frac{4(mh-b\lambda)}{7bm}]$$

$$W_{D^N/G/1/\infty} \approx W_1 + \frac{E(S)E(N)}{2} + \frac{Var(N) - E(N)}{E(N)}\frac{E(S)}{2}$$

$$= \frac{7b^2\lambda}{12(mh-b\lambda)} \cdot \exp\left[-\frac{4(mh-b\lambda)}{7bm}\right] + \frac{\frac{b\lambda}{m}}{2} + 0$$

$$= \frac{7b^2\lambda}{12(mh-b\lambda)} \cdot \exp\left[-\frac{4(mh-b\lambda)}{7bm}\right] + \frac{b\lambda}{2m}$$

$$W \le W_{D^N/G/1/\infty} \approx \frac{7b^2\lambda}{12(mh-b\lambda)} \cdot \exp\left[-\frac{4(mh-b\lambda)}{7bm}\right] + \frac{b\lambda}{2m} \qquad (2.2.1.T2.2)$$

Both the approximate upper bound and the strict upper bound are dimensionally correct. The strict upper bound is larger than the approximate upper bound.

Under heavy traffic,

$$\rho \to 1, mh - b\lambda \to 0, \exp\left[-\frac{4(mh-b\lambda)}{7bm}\right] \to 1,$$

the difference between the approximate and strict bounds is reduced to zero.

Note, as well, that in the limit, the ratio of the strict upper bound for the average waiting time under randomized assignment to the lower bound for the average waiting time:

$$\frac{Upper\ Bound}{Lower\ Bound} = \frac{\frac{7b^2\lambda}{12(mh-b\lambda)} + \frac{b\lambda}{2m}}{\frac{-7mbh + 7bh + 7b^2\lambda}{12(mh-b\lambda)}} = \frac{\lambda(7bm + 6hm - 6b\lambda)}{7m(h - hm + b\lambda)} = \frac{7\lambda bm}{7mh} \to m$$

when $\rho \to 1$.

2.2.2 Cyclic Assignment Policy

Another possible policy is to assign customers in cyclic order to the vehicles: the first customer in the batch is assigned to Vehicle 1, the second to Vehicle 2, ..., the *(m+1)-th* to Vehicle 1 again, and so forth. No jockeying of customers, after being assigned to vehicles, is allowed. Figure 6 illustrates this policy, which requires assigning an "identification number" to each vehicle to distinguish among them.

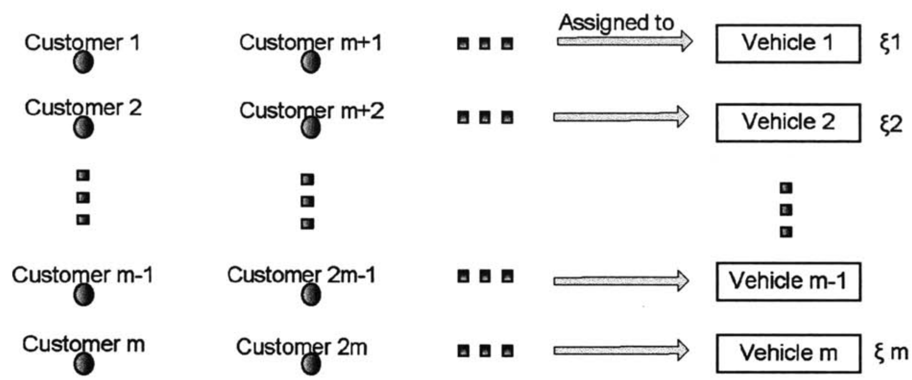


Figure 6: Cyclic assignment policy

The model corresponding to the cyclic assignment policy led (Technical Section 2.2.2.T1) to the following strict upper bound for the General distributions case:

$$W \le \frac{4mE^2(\xi)E(S^2) - 4E^2(S)E^3(\xi) + 4mhE(S)E(\xi^2) + m^3hE(S) - 4m^2hE(S)E(\xi)}{8m\big(mh - E(\xi)E(S)\big)E(\xi)} \qquad (7)$$

For Poisson batch sizes and a square service region, the bound (7) becomes (Technical Section 2.2.2.T2):

$$W \le \frac{14b^2\lambda^2m + 12b\lambda^2mh - 12b^2\lambda^3 + 12b\lambda mh - 12b\lambda m^2h + 3bm^3h}{24m\lambda(mh - b\lambda)} \qquad (8)$$

An approximate upper bound can also be obtained (Technical Section 2.2.2.T2) for the same case as (8):

$$W \leq \frac{(2m+12)b^2\lambda + 3b^2m^2}{24m(mh-b\lambda)} \cdot \exp\left[-\frac{8(mh-b\lambda)\lambda}{(2m+12)b\lambda + 3bm^2}\right] + \frac{4b\lambda^2 + 4b\lambda + bm^2 - 4b\lambda m}{8\lambda m} \quad (9)$$

A special case of (9) is also of interest in some applications. This is the case in which m/λ is large, i.e., the number of vehicles in the fleet is large relative to the rate at which customers arrive. This can be the situation during off-peak periods or when the vehicle fleet consists of a large pool of bicycles available for shared use. In such cases (9) becomes (Technical Section 2.2.2.T3):

$$W \leq \frac{7b^2\lambda m - 6b^2\lambda^2}{12m(mh-b\lambda)} \cdot \exp\left[-\frac{4(mh-b\lambda)}{7bm-6b\lambda}\right] \quad (10)$$

The approximate upper bound (10) has the desirable property of becoming more accurate as ρ approaches 1. Since $\rho = b\lambda/mh$, a large m/λ means a large b/h when ρ approaches 1. This corresponds to situations in which the service region is large and/or the train frequency is low.

2.2.2.T1*: Upper bound of the Unit-Capacity, Multi-Vehicle LMP under cyclic assignment policy in the general case

This policy consists of assigning customers in cyclic order to the m servers. After each batch arrival, the *1st* customer in the batch is assigned to the *1st* server, the *2nd* customer is assigned to the *2nd* server, ..., the *(m+1)-th* customer is assigned to the *1st* server again, and so forth. No jockeying of customers, after being assigned to vehicles, is allowed. We utilize different server orders for customers coming from different batches (trains).

There are totally m servers, with the name "Server *1*", "Server *2*",..., "Server m". Let N_i be the random variable indicating the number of customers assigned to "Server *i*" after the arrival of a particular train, with the assignment process upon arrival of each train being independent of the arrival process upon arrival of any other train.

After one train arrived, we order the m servers in sequence, the server receiving customers firstly is called "*1st* server", the server receiving customers secondly is called "*2nd* server", etc. Let X_i be the random variable indicating the number of customers assigned to the "$i\ th$ server" after the arrival of a particular train. Then,

$$X_1 = \left\lfloor \frac{\xi + m - 1}{m} \right\rfloor, X_2 = \left\lfloor \frac{\xi + m - 2}{m} \right\rfloor, \dots, X_{m-1} = \left\lfloor \frac{\xi + 1}{m} \right\rfloor, X_m = \left\lfloor \frac{\xi}{m} \right\rfloor$$

$$\xi = X_1 + X_2 + \cdots + X_{m-1} + X_m$$

The probability that Server i become the $j\ th$ server is $1/m$.

The modified model can be considered as $D^{N_i}/G/1/\infty$, which will provide an upper bound to the original $D^{\xi}/G/1/\infty$ model. $N_1, N_2, \dots, N_m$ are identically distributed, so all $D^{N_i}/G/1/\infty$ models can be taken as the same $D^N/G/1/\infty$ model, although $N_1, N_2, \dots, N_m$ are not necessarily independent. ξ is the random variable indicating the total number of customers coming from one train. $\xi = Km + R$, where $K = \lfloor \xi/m \rfloor$, and R is the remainder "left over" after division of ξ by m. So we can express ξ by a random vector with two dimension: (K, R).

$$X_i = \begin{cases} K+1, & 1 \le i \le R; \\ K, & R+1 \le i \le m; \end{cases}$$

$$E(N|(K,R)) = \frac{1}{m}[E(X_1|(K,R)) + E(X_2|(K,R)) + \dots + E(X_m|(K,R))]$$
$$= \frac{1}{m}[\left\lfloor \frac{\xi + m - 1}{m} \right\rfloor + \left\lfloor \frac{\xi + m - 2}{m} \right\rfloor + \left\lfloor \frac{\xi}{m} \right\rfloor] = \frac{Km + R}{m} = \frac{\xi}{m}$$

$$Var(N|(K,R)) = P(N = K+1) \cdot (K + 1 - E(N|(K,R)))^2 + P(N = K) \cdot (K - E(N|(K,R)))^2$$
$$= \sum_{i=1}^{R} P(N = X_i) \cdot (K + 1 - \frac{Km + R}{m})^2$$
$$+ \sum_{i=R+1}^{m} P(N = X_i) \cdot (K - \frac{Km + R}{m})^2 = \frac{R}{m} \cdot (K + 1 - \frac{Km + R}{m})^2 + \frac{m - R}{m}$$
$$\cdot (K - \frac{Km + R}{m})^2 = \frac{R}{m} \cdot \frac{(m - R)^2}{m^2} + \frac{m - R}{m} \cdot \frac{R^2}{m^2} = \frac{Rm - R^2}{m^2}$$

$$E(N) = E(E(N|(K,R))) = E(\frac{\xi}{m}) = \frac{E(\xi)}{m} \qquad (2.2.2.T1.1)$$

$$Var(N) = E(Var(N|(K,R))) + Var(E(N|(K,R))) = E(\frac{Rm - R^2}{m^2}) + Var(\frac{\xi}{m})$$

$$= \frac{E(Rm - R^2)}{m^2} + \frac{Var(\xi)}{m^2} \qquad (2.2.2.T1.2)$$

Since $Rm - R^2 \leq m^2/4$,

$$Var(N) = \frac{E(Rm - R^2)}{m^2} + \frac{Var(\xi)}{m^2} \leq \frac{1}{4} + \frac{Var(\xi)}{m^2} = \frac{4Var(\xi) + m^2}{4m^2} \qquad (2.2.2.T1.3)$$

Therefore,

$$W_{D^N/G/1/\infty} \leq \frac{E(N)Var(S) + E^2(S)Var(N)}{2(h - E(N)E(S))} + \frac{E(S)Var(N) + E(S)E^2(N) - E(S)E(N)}{2E(N)}$$

$$\leq \frac{\frac{E(\xi)}{m}Var(S) + E^2(S)\frac{4Var(\xi) + m^2}{4m^2}}{2\left(h - \frac{E(\xi)}{m}E(S)\right)} + \frac{E(S)\frac{4Var(\xi) + m^2}{4m^2} + E(S)\left(\frac{E(\xi)}{m}\right)^2 - E(S)\frac{E(\xi)}{m}}{2\frac{E(\xi)}{m}}$$

$$\leq \frac{4mE(\xi)Var(S) + m^2E^2(S) + 4E^2(S)Var(\xi)}{8m(mh - E(\xi)E(S))}$$

$$+ \frac{4E(S)Var(\xi) + E(S)m^2 + 4E(S)E^2(\xi) - 4mE(S)E(\xi)}{8mE(\xi)}$$

$$= \frac{4mE(\xi)E(S^2) - 4mE(\xi)E^2(S) + m^2E^2(S) + 4E^2(S)E(\xi^2) - 4E^2(S)E^2(\xi)}{8m(mh - E(\xi)E(S))}$$

$$+ \frac{E(S)(4E(\xi^2) + m^2 - 4mE(\xi))}{8mE(\xi)}$$

$$= \frac{4mE^2(\xi)E(S^2) - 4E^2(S)E^3(\xi) + 4mhE(S)E(\xi^2) + m^3hE(S) - 4m^2hE(S)E(\xi)}{8m(mh - E(\xi)E(S))E(\xi)}$$

This is a strict upper bound for the average waiting time under cyclic assignment policy in the general case.

2.2.2.T2*: Upper bound of the Unit-Capacity, Multi-Vehicle LMP under cyclic assignment policy for a Poisson customer batch size and a square service region

If the number of customers from each train is Poisson-distributed and the service time is as the square service region case described in Technical Section 2.1.T2, under cyclic assignment policy, we know:

$E(\xi) = \lambda$

$Var(\xi) = \lambda$

$E(\xi^2) = Var(\xi) + E^2(\xi) = \lambda + \lambda^2$

$E(S) = b$

$Var(S) = \frac{b^2}{6}$

$E(S^2) = Var(S) + E^2(S) = \frac{7b^2}{6}$

$E(N) = \frac{\lambda}{m}$

$Var(N) \leq \frac{4Var(\xi) + m^2}{4m^2} = \frac{4\lambda + m^2}{4m^2}$

$C_a^2 = 0$ (Due to constant batch or macro customer inter-arrival time)

$$C_s^2 = C_t^2 = \frac{E(N)Var(S) + E^2(S)Var(N)}{E^2(N)E^2(S)} = \frac{\frac{\lambda}{m}\frac{b^2}{6} + b^2\frac{4\lambda + m^2}{4m^2}}{\frac{\lambda^2}{m^2}b^2} = \frac{3m^2 + 12\lambda + 2m\lambda}{12\lambda^2}$$

$$s = E(T) = E(N)E(S) = \frac{b\lambda}{m}$$

$$\rho = \frac{E(T)}{h} = \frac{E(N)E(S)}{h} = \frac{b\lambda}{mh}$$

Thus, we can obtain a strict upper bound for the average waiting time in the original $D^\xi/G/m/\infty$ model:

$$W \leq \frac{4mE^2(\xi)E(S^2) - 4E^2(S)E^3(\xi) + 4mhE(S)E(\xi^2) + m^3hE(S) - 4m^2hE(S)E(\xi)}{8m\big(mh - E(\xi)E(S)\big)E(\xi)}$$

$$= \frac{4m\lambda^2\frac{7b^2}{6} - 4b^2\lambda^3 + 4mhb(\lambda + \lambda^2) + m^3hb - 4m^2hb\lambda}{8m(mh - \lambda b)\lambda}$$

$$= \frac{14b^2\lambda^2m + 12b\lambda^2mh - 12b^2\lambda^3 + 12b\lambda mh - 12b\lambda m^2h + 3bm^3h}{24m\lambda(mh - b\lambda)}$$

Similarly, we have the approximation (using the same notations as in Technical Section 2.2.T):

$$W_1 \approx \frac{\rho}{1-\rho} \cdot \frac{(C_a^2 + C_s^2)}{2} s \cdot \exp\left[-\frac{2(1-\rho)(1-C_a^2)^2}{3\rho(C_a^2 + C_s^2)}\right]$$

$$= \frac{\frac{b\lambda}{mh}}{1-\frac{b\lambda}{mh}} \cdot \frac{\frac{3m^2+12\lambda+2m\lambda}{12\lambda^2}}{2} \frac{b\lambda}{m} \cdot \exp\left[-\frac{2\left(1-\frac{b\lambda}{mh}\right)}{3\frac{b\lambda}{mh}\frac{3m^2+12\lambda+2m\lambda}{12\lambda^2}}\right]$$

$$= \frac{(2m+12)b^2\lambda + 3b^2m^2}{24m(mh-b\lambda)} \cdot \exp\left[-\frac{8(mh-b\lambda)\lambda}{(2m+12)b\lambda+3bm^2}\right]$$

$$W_{D^N/G/1/\infty} = W_1 + \frac{E(S)E(N)}{2} + \frac{Var(N)-E(N)}{E(N)}\frac{E(S)}{2}$$

$$\approx \frac{(2m+12)b^2\lambda + 3b^2m^2}{24m(mh-b\lambda)} \cdot \exp\left[-\frac{8(mh-b\lambda)\lambda}{(2m+12)b\lambda+3bm^2}\right] + \frac{\frac{b\lambda}{m}}{2} + \frac{\frac{4\lambda+m^2}{4m^2} - \frac{\lambda}{m}}{\frac{\lambda}{m}}\frac{b}{2}$$

$$= \frac{(2m+12)b^2\lambda + 3b^2m^2}{24m(mh-b\lambda)} \cdot \exp\left[-\frac{8(mh-b\lambda)\lambda}{(2m+12)b\lambda+3bm^2}\right] + \frac{4b\lambda^2 + 4b\lambda + bm^2 - 4b\lambda m}{8\lambda m}$$

$$W \le W_{D^N/G/1/\infty}$$

$$\approx \frac{(2m+12)b^2\lambda + 3b^2m^2}{24m(mh-b\lambda)} \cdot \exp\left[-\frac{8(mh-b\lambda)\lambda}{(2m+12)b\lambda+3bm^2}\right] + \frac{4b\lambda^2 + 4b\lambda + bm^2 - 4b\lambda m}{8\lambda m} \quad (2.2.2.T2.2)$$

Both the approximate upper bound and the strict upper bounds are correct dimensionally. The strict upper bound is a little larger than the approximate upper bound.

Under heavy traffic,

$$\rho \to 1, mh - b\lambda \to 0, \exp\left[-\frac{8(mh-b\lambda)\lambda}{(2m+12)b\lambda+3bm^2}\right] \to 1,$$

the difference between the approximate and strict bounds decreases to zero.

Note, as well, that in the limit, the ratio of the strict upper bound for the average waiting time under cyclic assignment to the lower bound for the average waiting time:

$$\frac{Upper\ Bound}{Lower\ Bound}=\frac{\frac{(2m+12)b^2\lambda+3b^2m^2}{24m(mh-b\lambda)}+\frac{4b\lambda^2+4b\lambda+bm^2-4b\lambda m}{8\lambda m}}{\frac{-7mbh+7bh+7b^2\lambda}{12(mh-b\lambda)}}$$

$$=-\frac{(hm-b\lambda)(bm\lambda(3m^2+12\lambda+2m\lambda)+3(m^2-4m\lambda+4\lambda(1+\lambda))m(hm-b\lambda))}{14m\lambda(h(-1+m)-b\lambda)m(hm-b\lambda)}$$

$$=\frac{2b\lambda^2(-7m+6\lambda)-3hm(m^2-4m\lambda+4\lambda(1+\lambda))}{14m\lambda(h(-1+m)-b\lambda)}$$

$$=\frac{2\lambda(-7m+6\lambda)-3(m^2-4m\lambda+4\lambda(1+\lambda))}{-14\lambda}\rightarrow\frac{1}{14}(12+2m+\frac{3m^2}{\lambda})$$

when $\rho \rightarrow 1$.

2.2.2.T3*: Upper bound of the Unit-Capacity, Multi-Vehicle LMP under cyclic assignment policy for a Poisson customer batch size and a square service region when m/λ is large

$E(\xi)=\lambda$

$Var(\xi)=\lambda$

$E(\xi^2)=Var(\xi)+E^2(\xi)=\lambda+\lambda^2$

$E(S)=b$

$Var(S)=\frac{b^2}{6}$

$E(S^2)=Var(S)+E^2(S)=\frac{7b^2}{6}$

$C_a^2=0$ (Due to constant batch or macro customer inter-arrival time)

$E(N)=\frac{\lambda}{m}$

When m/λ is large, because $E(Rm - R^2) \approx m^2/4$ is not a good approximation, we need to obtain a better approximation.

According to numerical experiment, when m/λ is large, we obtain:

$E(Rm - R^2) \to m\lambda - \lambda(\lambda + 1)$

Therefore,

$$Var(N) = E(Var(N|(K,R))) + Var(E(N|(K,R))) = E(\frac{Rm - R^2}{m^2}) + Var(\frac{\xi}{m})$$
$$= \frac{E(Rm - R^2)}{m^2} + \frac{\lambda}{m^2} \approx \frac{m\lambda - \lambda(\lambda + 1)}{m^2} + \frac{\lambda}{m^2} = \frac{m\lambda - \lambda^2}{m^2}$$

$$C_s^2 = C_t^2 = \frac{E(N)Var(S) + E^2(S)Var(N)}{E^2(N)E^2(S)} = \frac{\frac{\lambda}{m}\frac{b^2}{6} + b^2\frac{m\lambda - \lambda^2}{m^2}}{\frac{\lambda^2}{m^2}b^2} = -1 + \frac{7m}{6\lambda}$$

$$s = E(T) = E(N)E(S) = \frac{b\lambda}{m}$$

$$\rho = \frac{E(T)}{h} = \frac{E(N)E(S)}{h} = \frac{b\lambda}{mh}$$

From the same analysis, we obtain a strict upper bound for the average waiting time in the original $D^{\xi}/G/m/\infty$ model when m/λ is large:

$$W \leq \frac{E(N)Var(S) + E^2(S)Var(N)}{2(h - E(N)E(S))} + \frac{E(S)Var(N) + E(S)E^2(N) - E(S)E(N)}{2E(N)}$$
$$= \frac{\frac{\lambda}{m}\frac{b^2}{6} + b^2\frac{m\lambda - \lambda^2}{m^2}}{2\left(h - \frac{\lambda}{m}b\right)} + \frac{b\frac{m\lambda - \lambda^2}{m^2} + b\frac{\lambda^2}{m^2} - b\frac{\lambda}{m}}{2\frac{\lambda}{m}} = \frac{7b^2\lambda m - 6b^2\lambda^2}{12m(mh - b\lambda)}$$

Similarly, we have the approximation (using the same notations as in Technical Section 2.2.T):

$$W_1 \approx \frac{\rho}{1-\rho} \cdot \frac{(C_a^2 + C_s^2)}{2} s \cdot \exp\left[-\frac{2(1-\rho)(1-C_a^2)^2}{3\rho(C_a^2+C_s^2)}\right]$$

$$= \frac{\frac{b\lambda}{mh}}{1-\frac{b\lambda}{mh}} \cdot \frac{-1+\frac{7m}{6\lambda}}{2} \cdot \frac{b\lambda}{m} \cdot \exp\left[-\frac{2\left(1-\frac{b\lambda}{mh}\right)}{3\frac{b\lambda}{mh}(-1+\frac{7m}{6\lambda})}\right]$$

$$= \frac{7b^2\lambda m - 6b^2\lambda^2}{12m(mh-b\lambda)} \cdot \exp\,[-\frac{4(mh-b\lambda)}{7bm-6b\lambda}]$$

$$W_{D^N/G/1/\infty} = W_1 + \frac{E(S)E(N)}{2} + \frac{Var(N)-E(N)}{E(N)}\frac{E(S)}{2}$$

$$\approx \frac{7b^2\lambda m - 6b^2\lambda^2}{12m(mh-b\lambda)} \cdot \exp\,[-\frac{4(mh-b\lambda)}{7bm-6b\lambda}] + \frac{\frac{b\lambda}{m}}{2} + \frac{\frac{m\lambda-\lambda^2}{m^2}-\frac{\lambda}{m}}{\frac{\lambda}{m}}\frac{b}{2}$$

$$= \frac{7b^2\lambda m - 6b^2\lambda^2}{12m(mh-b\lambda)} \cdot \exp\,[-\frac{4(mh-b\lambda)}{7bm-6b\lambda}]$$

$$W \le W_{D^N/G/1/\infty} \approx \frac{7b^2\lambda m - 6b^2\lambda^2}{12m(mh-b\lambda)} \cdot \exp\,[-\frac{4(mh-b\lambda)}{7bm-6b\lambda}] \qquad (2.2.2.T3.2)$$

Both the approximate upper bound and the strict upper bound are correct dimensionally. The strict upper bound is a little larger than the approximate upper bound.

Under heavy traffic,

$$\rho \to 1, mh - b\lambda \to 0, \exp\left[\,[-\frac{4(mh-b\lambda)}{7bm-6b\lambda}\right] \to 1,$$

the difference between the approximate and strict bounds decreases to zero.

Note, as well, that in the limit, the ratio of the strict upper bound for the average waiting time under cyclic assignment to the lower bound for the average waiting time:

$$\frac{Upper\ Bound}{Lower\ Bound} = \frac{\frac{7b^2\lambda m - 6b^2\lambda^2}{12m(mh-b\lambda)}}{\frac{-7mbh+7bh+7b^2\lambda}{12(mh-b\lambda)}} = -\frac{b(7m-6\lambda)\lambda(-hm+b\lambda)}{7(h-hm+b\lambda)m(hm-b\lambda)} =$$

$$= \frac{b(7m-6\lambda)\lambda}{7(h-hm+b\lambda)m} \to m - \frac{6\lambda}{7}$$

when $\rho \to 1$ and m/λ is large.

2.3 Numerical Experiments for the Unit-Capacity, Multi-Vehicle LMP

To assess the performance of the many approximate expressions obtained in Chapter 2.1 and 2.2 under a broad range of conditions, a simple simulation of the Unit-Capacity, Multi-Vehicle LMP was carried out with a program written in java. We consider a square service region with geometry $a/v_x = b/v_y = 2.5\, min = 150\, sec$, headway of $h = 10\, min = 600\, sec$, and Poisson-distributed batch sizes of $\lambda = 20, 40, 60, 80$. We selected these parameters so that the system would make sense physically. The respective simulation results are shown in Figures 7, 8, 9, and 10.

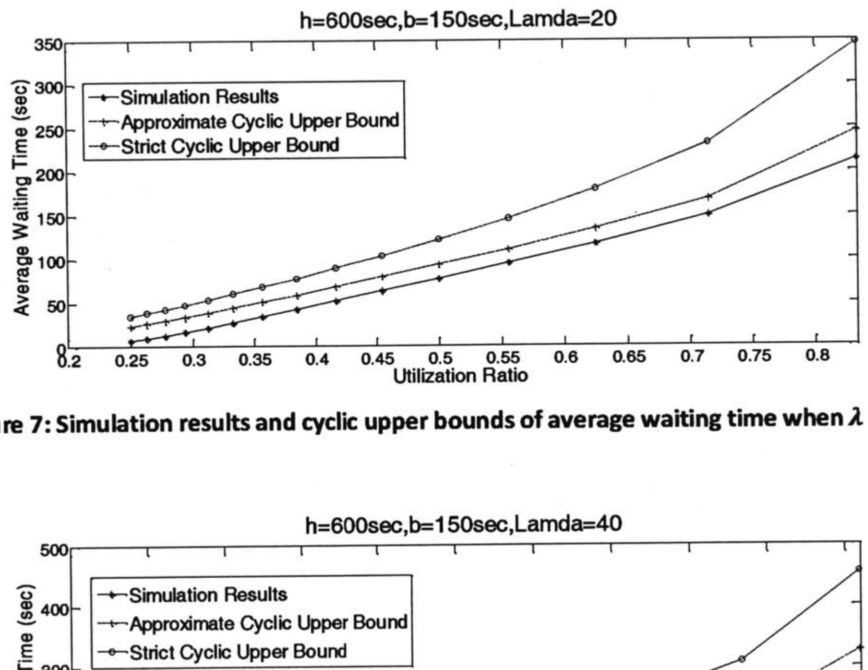


Figure 7: Simulation results and cyclic upper bounds of average waiting time when $\lambda = 20$

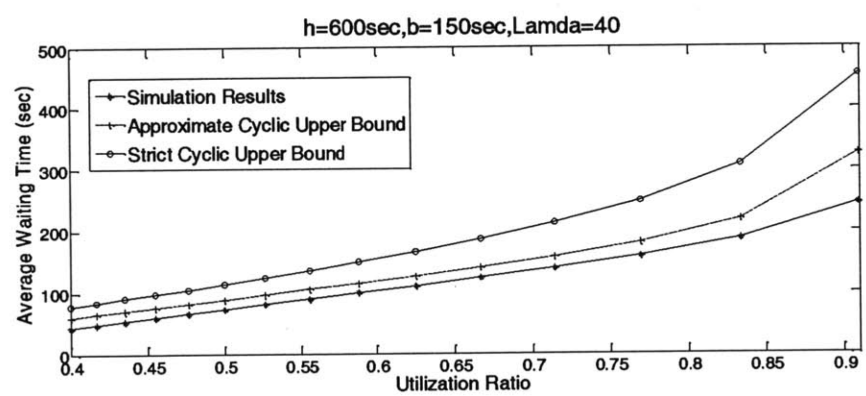


Figure 8: Simulation results and cyclic upper bounds of average waiting time when $\lambda = 40$

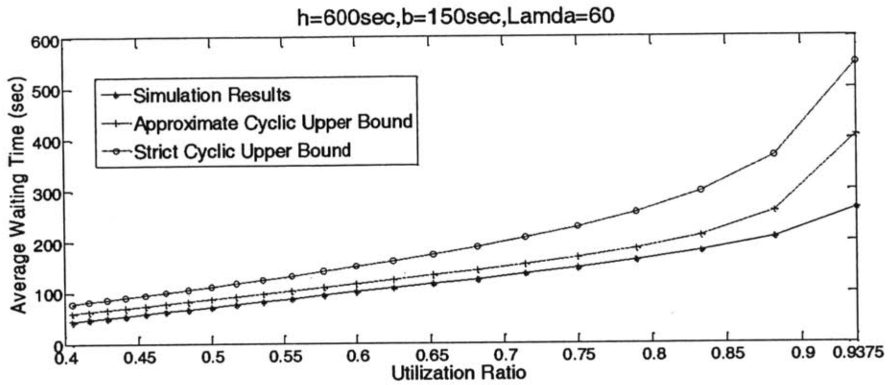


Figure 9: Simulation results and cyclic upper bounds of average waiting time when $\lambda = 60$

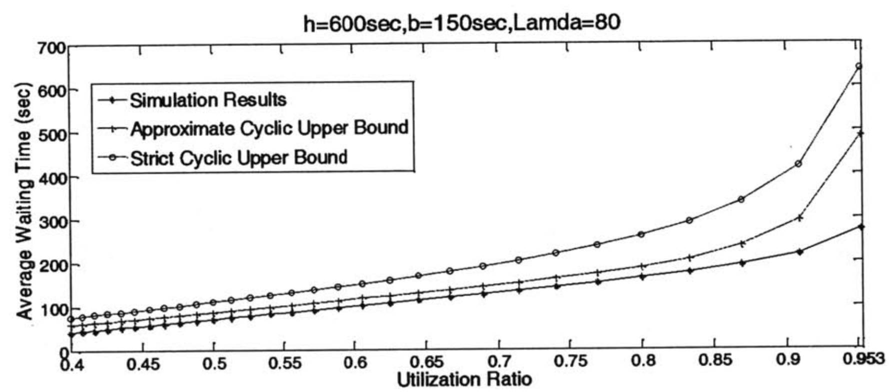


Figure 10: Simulation results and cyclic upper bounds of average waiting time when $\lambda = 80$.

The figures plot the simulation results and our estimates for the average waiting time per customer W (in seconds) against the utilization ratio = $b\lambda/mh$. Since the simulated system has Poisson customer batch size and a square service region, and m/λ is not large, only expressions (2), (5), (6), (8), and (9) from Chapter 2.1 and 2.2 are applicable and considered here.

Comparison with the simulation results led to two initial observations: first, the strict lower bound (2) is not useful, as it provides poor estimates of W, often including negative values; and, second, the strict randomized assignment upper bound (5) and the approximate randomized assignment upper bound (6) is also unreliable as it often generates very high estimates of delays. The values obtained from (5) and (6) have therefore been omitted from Figures 7-10, which only show the strict cyclic upper bound (8), the approximate cyclic upper bound (9) and the simulation results.

As can be seen in the figures, the strict cyclic upper bound, (8), is a consistently reliable upper bound for W, while the approximate cyclic upper bound, (9), provides a very good approximation for the entire range of parameter values explored, which span the full set of conditions under which the LMTS remains stable. In a practical system, it would be desirable to achieve values of 1 to 5 minutes, for the average waiting time until customers to board a vehicle. Note from Figures 7-10 that for this range of values (60 to 300 seconds) the difference between the approximate cyclic upper bound and the simulation results stays small in both absolute and percentage terms. For example, when $\lambda = 20$ (Figure 7), this difference never exceeds 30 seconds and 15% for values of W between 2 and 4 minutes. For a queuing system as analytically complicated as $D^{\xi}/G/m/\infty$, expression (9) performs remarkably well.

We also note that it is not surprising that (9), the approximate cyclic upper bound, performs much better than (6), the approximate randomized upper bound. This is because the customers are more evenly distributed among the vehicles under the cyclic assignment policy than under the randomized assignment policy and, consequently, the variance of the service times under the former policy is much smaller than under the latter for instances of practical interest.

In conclusion, given the train frequency (batch inter-arrival times), customer arrival intensity (batch size), geometry of the service region (shape and size), distance metric (right-angle, Euclidean) and vehicle speed, we can use expressions based on the strict cyclic upper bound, (8) and the approximate cyclic upper bound, (9), to estimate LMTS system performance for any given number of unit-capacity vehicles. Chapter 2.4 will first demonstrate the robustness of (8) and (9) to mild changes in the assumptions under which they were obtained. In Chapter 3, we shall seek to extend our findings to the general case in which vehicle capacity can be greater than 1.

2.4 Sensitivity Analysis: Unit-Capacity, Multi-Vehicle LMP

In this section, we relax the assumptions concerning the shape of the service region and the continuity of the travel medium to derive expressions for W, analogous to (2), (5), (6), (8), and (9), for three specific cases: a rectangular service region; a diamond-shaped region; and a service region that includes a barrier to travel. We then repeat our simulation experiments to test the performance of the new expressions and conclude that the strict cyclic upper bound and the approximate cyclic upper bound continue to outperform the other bounds and to provide accurate approximations to W under a wide range of conditions.

2.4.1 Rectangular Service Region ($a = kb, k > 1$)

The service region is now assumed to be a rectangle with length of a and width of b, as illustrated in Figure 11. Travel is according to the right-angle metric in directions parallel to the sides of the rectangle.

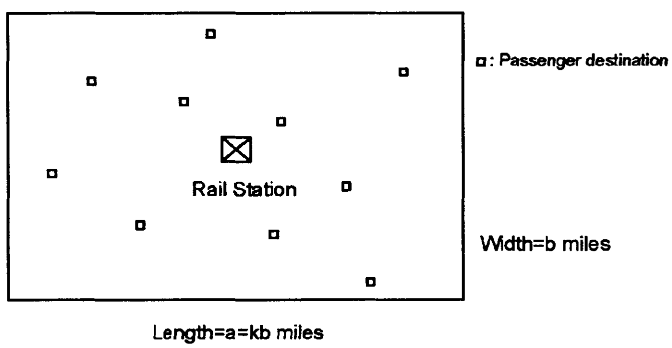


Figure 11: Rectangular service region

The expressions for the five strict and approximate bounds for this case are derived in Technical Section 2.4.1.T. For the simulation experiment, we considered two examples:

(i) $a/v_x = 3\text{ min} = 180\text{ sec}, b/v_y = 2\text{ min} = 120\text{ sec}$;

(ii) $a/v_x = 4 \text{ min} = 240 \text{ sec}, b/v_y = 2 \text{ min} = 120 \text{ sec}$;

The headway h is set at 600 sec and the batch size of arriving customers at the train station is assumed to be Poisson-distributed with $\lambda = 20, 40, 60, 80$.

A typical instance of the results and comparisons for just one case (Example (i) with $\lambda = 20$) is shown in Figure 12. As in Figures 7-10, the theoretical estimates shown are limited to those obtained through the best performing expressions, namely the strict cyclic upper bound and the approximate cyclic upper bound.

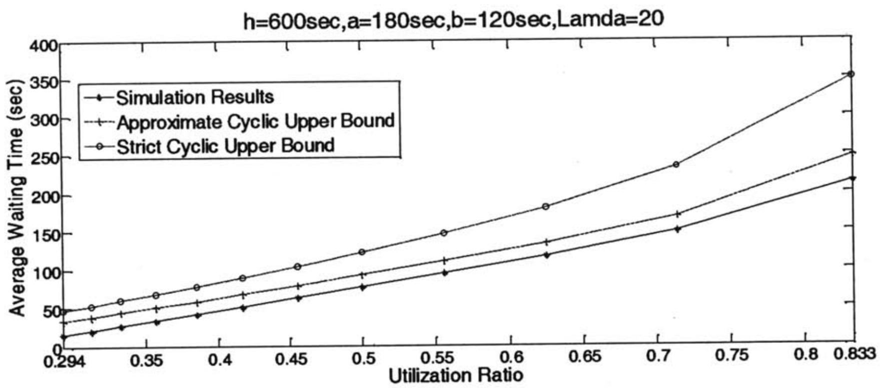


Figure 12: Simulation results and cyclic upper bounds when $a = 180sec, b = 120sec, \lambda = 20$

For Example (i), i.e., for $k = 1.5$, and for values of the average waiting time of the order of 1 to 4 minutes, the percent difference between the approximate cyclic upper bound and the simulation results is of the order of 10-25% for the entire range of values of λ (= 20, 40, 60, 80). For Example (ii), i.e., for $k = 2$, $a = 240sec$, $b = 120sec$, this increased to 20-35%. Thus, as k becomes larger and the service region more elongated, the approximate cyclic upper bound becomes less accurate. This is because this approximate bound is sensitive to the variance of the service times which, in turn, increases as the region becomes more elongated and resembles a rectangular strip. The bound's accuracy is, however, relatively insensitive to the customer demand intensity λ.

2.4.1.T*: Sensitivity analysis of Unit-Capacity, Multi-Vehicle LMP for rectangle service region with length a and width b ($a = kb, k > 1$)

S is individual customer service time (round trip) in the region, we know:

$$E(S) = \frac{(k+1)b}{2}, Var(S) = \frac{(k^2+1)b^2}{12},$$

$$E(S^2) = \frac{(2k^2+2+3k)b^2}{6}, C_s^2 = \frac{k^2+1}{3(k+1)^2},$$

$$\rho = \frac{E(T)}{h} = \frac{E(N)E(S)}{h} = \frac{(k+1)b\lambda}{2mh}$$

$$s = E(S)E(N) = \frac{\frac{(k+1)b}{2}\lambda}{m} = \frac{(k+1)b\lambda}{2m}$$

Strict Lower Bound:

$$W \geq \frac{E(\xi)E(S)E(S^2) + hE(S^2) - 2hE^2(S) - mhE(S^2)}{2E(S)\left(mh - E(\xi)E(S)\right)}$$

$$\geq \frac{\lambda\frac{(k+1)b}{2}\frac{(2k^2+2+3k)b^2}{6} + h\frac{(2k^2+2+3k)b^2}{6} - 2h\frac{(k+1)^2b^2}{4} - mh\frac{(2k^2+2+3k)b^2}{6}}{2\frac{(k+1)b}{2}\left(mh - \lambda\frac{(k+1)b}{2}\right)}$$

$$= -\frac{b(-2h(1+2m+3k(1+m)+k^2(1+2m)) + b(2+5k+5k^2+2k^3)\lambda)}{6(1+k)(-2hm+b(1+k)\lambda)}$$

$$= \frac{b(-2h(1+2m+3k(1+m)+k^2(1+2m)) + b(2+5k+5k^2+2k^3)\lambda)}{6(1+k)(2hm-b(1+k)\lambda)} \qquad (2.4.1.T.1)$$

Randomized Upper Bound:

$$E(\xi) = \lambda$$

$$Var(\xi) = \lambda$$

$$E(\xi^2) = Var(\xi) + E^2(\xi) = \lambda + \lambda^2$$

$$E(S) = \frac{(k+1)b}{2}$$

$$Var(S) = \frac{(k^2+1)b^2}{12}$$

$$E(S^2) = Var(S) + E^2(S) = \frac{(2k^2+2+3k)b^2}{6}$$

$$E(N) = \frac{\lambda}{m}$$

$$Var(N) = \frac{\lambda}{m}$$

$C_a^2 = 0$ (Due to constant batch or macro customer inter-arrival time)

$$C_s^2 = C_t^2 = \frac{E(N)Var(S) + E^2(S)Var(N)}{E^2(N)E^2(S)} = \frac{\frac{\lambda}{m}\frac{(k^2+1)b^2}{12} + \frac{(k+1)^2b^2}{4}\frac{\lambda}{m}}{\frac{\lambda^2}{m^2}\frac{(k+1)^2b^2}{4}}$$

$$= \frac{2(2+3k+2k^2)m}{3(1+k)^2\lambda}$$

$$s = E(S)E(N) = \frac{\frac{(k+1)b}{2}\lambda}{m} = \frac{(k+1)b\lambda}{2m}$$

$$\rho = \frac{E(T)}{h} = \frac{E(N)E(S)}{h} = \frac{(k+1)b\lambda}{2mh}$$

Thus, using the general conclusion, we can obtain a strict upper bound for the average waiting time in the original $D^{\xi}/G/m/\infty$ model:

$$W \le \frac{mhE(S)E(\xi^2) - mhE(S)E(\xi) + mE(S^2)E^2(\xi) - E^2(S)E^3(\xi)}{2m\big(mh - E(\xi)E(S)\big)E(\xi)}$$

$$= \frac{mh\frac{(k+1)b}{2}(\lambda+\lambda^2) - mh\frac{(k+1)b}{2}\lambda + m\frac{(2k^2+2+3k)b^2}{6}\lambda^2 - \frac{(k+1)^2b^2}{4}\lambda^3}{2m\left(mh - \lambda\frac{(k+1)b}{2}\right)\lambda}$$

$$= \frac{b\lambda(6h(1+k)m + b((4+6k+4k^2)m - 3(1+k)^2\lambda))}{12m(2mh - b(1+k)\lambda)} \qquad (2.4.1.T.2)$$

Similarly, we have the approximation:

$$W_1 \approx \frac{\rho}{1-\rho} \cdot \frac{(C_a^2+C_s^2)}{2} s \cdot \exp\left[-\frac{2(1-\rho)(1-C_a^2)^2}{3\rho(C_a^2+C_s^2)}\right]$$

$$= \frac{\frac{(k+1)b\lambda}{2mh}}{1-\frac{(k+1)b\lambda}{2mh}} \cdot \frac{\frac{2(2+3k+2k^2)m}{3(1+k)^2\lambda}}{2} \frac{(k+1)b\lambda}{2m} \cdot \exp\left[-\frac{2\left(1-\frac{(k+1)b\lambda}{2mh}\right)}{3\frac{(k+1)b\lambda}{2mh}\frac{2(2+3k+2k^2)m}{3(1+k)^2\lambda}}\right]$$

$$= \frac{b^2(2+3k+2k^2)\lambda}{6(2mh-b(1+k)\lambda)} \cdot \exp\left[\frac{(1+k)(-2hm+b(1+k)\lambda)}{b(2+3k+2k^2)m}\right]$$

$$W_{D^N/G/1/\infty} = W_1 + \frac{E(S)Var(N)+E(S)E^2(N)-E(S)E(N)}{2E(N)}$$

$$\approx \frac{b^2(2+3k+2k^2)\lambda}{6(2mh-b(1+k)\lambda)} \cdot \exp\left[\frac{(1+k)(-2hm+b(1+k)\lambda)}{b(2+3k+2k^2)m}\right] + \frac{\frac{(k+1)b}{2}\frac{\lambda}{m}+\frac{(k+1)b}{2}\frac{\lambda^2}{m^2}-\frac{(k+1)b}{2}\frac{\lambda}{m}}{2\frac{\lambda}{m}}$$

$$= \frac{b^2(2+3k+2k^2)\lambda}{6(2mh-b(1+k)\lambda)} \cdot \exp\left[\frac{(1+k)(-2hm+b(1+k)\lambda)}{b(2+3k+2k^2)m}\right] + \frac{b(1+k)\lambda}{4m}$$

$$W \le W_{D^N/G/1/\infty} \approx \frac{b^2(2+3k+2k^2)\lambda}{6(2mh-b(1+k)\lambda)} \cdot \exp\left[\frac{(1+k)(-2hm+b(1+k)\lambda)}{b(2+3k+2k^2)m}\right] + \frac{b(1+k)\lambda}{4m}$$

Both the approximate upper bound and the strict upper bound are dimensionally correct. The strict upper bound is larger than the approximate upper bound.

Under heavy traffic,

$$\rho \to 1, 2mh-(k+1)b\lambda \to 0, \exp\left[\frac{(1+k)(-2hm+b(1+k)\lambda)}{b(2+3k+2k^2)m}\right] \to 1,$$

the difference between the approximate and strict bounds is reduced to zero.

Cyclic Upper Bound:

$E(\xi) = \lambda$

$$Var(\xi) = \lambda$$

$$E(\xi^2) = Var(\xi) + E^2(\xi) = \lambda + \lambda^2$$

$$E(S) = \frac{(k+1)b}{2}$$

$$Var(S) = \frac{(k^2+1)b^2}{12}$$

$$E(S^2) = Var(S) + E^2(S) = \frac{(2k^2+2+3k)b^2}{6}$$

$$E(N) = \frac{\lambda}{m}$$

$$Var(N) \leq \frac{4Var(\xi) + m^2}{4m^2} = \frac{4\lambda + m^2}{4m^2}$$

$C_a^2 = 0$ (Due to constant batch or macro customer inter-arrival time)

$$C_s^2 = C_t^2 = \frac{E(N)Var(S) + E^2(S)Var(N)}{E^2(N)E^2(S)} \leq \frac{\frac{\lambda}{m}\frac{(k^2+1)b^2}{12} + \frac{(k+1)^2b^2}{4}\frac{4\lambda+m^2}{4m^2}}{\frac{\lambda^2}{m^2}\frac{(k+1)^2b^2}{4}}$$
$$= \frac{3(1+k)^2m^2 + 12(1+k)^2\lambda + 4(1+k^2)m\lambda}{12(1+k)^2\lambda^2}$$

$$s = E(S)E(N) = \frac{\frac{(k+1)b}{2}\lambda}{m} = \frac{(k+1)b\lambda}{2m}$$

$$\rho = \frac{E(T)}{h} = \frac{E(N)E(S)}{h} = \frac{(k+1)b\lambda}{2mh}$$

Thus, using the general conclusion, we can obtain a strict upper bound for the average waiting time in the original $D^{\xi}/G/m/\infty$ model:

$$W \le \frac{4mE^2(\xi)E(S^2) - 4E^2(S)E^3(\xi) + 4mhE(S)E(\xi^2) + m^3hE(S) - 4m^2hE(S)E(\xi)}{8m\big(mh - E(\xi)E(S)\big)E(\xi)}$$

$$= \frac{4m\lambda^2\frac{(2k^2+2+3k)b^2}{6} - 4\frac{(k+1)^2b^2}{4}\lambda^3 + 4mh\frac{(k+1)b}{2}(\lambda+\lambda^2) + m^3h\frac{(k+1)b}{2} - 4m^2h\frac{(k+1)b}{2}\lambda}{8m\left(mh - \lambda\frac{(k+1)b}{2}\right)\lambda}$$

$$= \frac{b(2b\lambda^2((4+6k+4k^2)m - 3(1+k)^2\lambda) + 3h(1+k)m(m^2 - 4m\lambda + 4\lambda(1+\lambda)))}{24m\lambda(2hm - b(1+k)\lambda)} \qquad (2.4.1.T.4)$$

Similarly, we have the approximation:

$$W_1 \approx \frac{\rho}{1-\rho}\cdot\frac{(C_a^2 + C_s^2)}{2}s\cdot\exp\left[-\frac{2(1-\rho)(1-C_a^2)^2}{3\rho(C_a^2+C_s^2)}\right]$$

$$= \frac{\frac{(k+1)b\lambda}{2mh}}{1-\frac{(k+1)b\lambda}{2mh}}\,\frac{\frac{3(1+k)^2m^2 + 12(1+k)^2\lambda + 4(1+k^2)m\lambda}{12(1+k)^2\lambda^2}}{2}\,\frac{(k+1)b\lambda}{2m}$$

$$\cdot\exp\left[-\frac{2\left(1-\frac{(k+1)b\lambda}{2mh}\right)}{3\frac{(k+1)b\lambda}{2mh}\frac{3(1+k)^2m^2 + 12(1+k)^2\lambda + 4(1+k^2)m\lambda}{12(1+k)^2\lambda^2}}\right]$$

$$= \frac{b^2(3(1+k)^2m^2 + 12(1+k)^2\lambda + 4(1+k^2)m\lambda)}{48m(2hm - b(1+k)\lambda)}$$

$$\cdot\exp\left[\frac{8(1+k)\lambda(-2hm + b(1+k)\lambda)}{b(3(1+k)^2m^2 + 12(1+k)^2\lambda + 4(1+k^2)m\lambda)}\right]$$

$$W_{D^N/G/1/\infty} = W_1 + \frac{E(S)Var(N) + E(S)E^2(N) - E(S)E(N)}{2E(N)}$$

$$\leq \frac{b^2(3(1+k)^2m^2 + 12(1+k)^2\lambda + 4(1+k^2)m\lambda)}{48m(2hm - b(1+k)\lambda)}$$

$$\cdot \exp\left[\frac{8(1+k)\lambda(-2hm + b(1+k)\lambda)}{b(3(1+k)^2m^2 + 12(1+k)^2\lambda + 4(1+k^2)m\lambda)}\right]$$

$$+ \frac{\frac{(k+1)b}{2}\frac{4\lambda + m^2}{4m^2} + \frac{(k+1)b}{2}\frac{\lambda^2}{m^2} - \frac{(k+1)b}{2}\frac{\lambda}{m}}{2\frac{\lambda}{m}}$$

$$= \frac{b^2(3(1+k)^2m^2 + 12(1+k)^2\lambda + 4(1+k^2)m\lambda)}{48m(2hm - b(1+k)\lambda)}$$

$$\cdot \exp\left[\frac{8(1+k)\lambda(-2hm + b(1+k)\lambda)}{b(3(1+k)^2m^2 + 12(1+k)^2\lambda + 4(1+k^2)m\lambda)}\right]$$

$$+ \frac{b(1+k)(m^2 - 4m\lambda + 4\lambda(1+\lambda))}{16m\lambda}$$

$$W \leq W_{D^N/G/1/\infty}$$

$$\leq \frac{b^2(3(1+k)^2m^2 + 12(1+k)^2\lambda + 4(1+k^2)m\lambda)}{48m(2hm - b(1+k)\lambda)}$$

$$\cdot \exp\left[\frac{8(1+k)\lambda(-2hm + b(1+k)\lambda)}{b(3(1+k)^2m^2 + 12(1+k)^2\lambda + 4(1+k^2)m\lambda)}\right]$$

$$+ \frac{b(1+k)(m^2 - 4m\lambda + 4\lambda(1+\lambda))}{16m\lambda} \qquad (2.4.1.T.5)$$

Both the approximate upper bound and the strict upper bound are dimensionally correct. The strict upper bound is larger than the approximate upper bound.

Under heavy traffic,

$$\rho \to 1, 2mh - (k+1)b\lambda \to 0, \exp\left[\frac{8(1+k)\lambda(-2hm + b(1+k)\lambda)}{b(3(1+k)^2m^2 + 12(1+k)^2\lambda + 4(1+k^2)m\lambda)}\right] \to 1,$$

the difference between the approximate and strict bounds is reduced to zero.

2.4.2 Diamond Service Region with Side of Length *b*

In the next sensitivity test, the service region is assumed to be a perfect four-sided diamond with side equal to b, as illustrated in Figure 13. The theoretical results for this case are derived in Technical Section 2.4.2.T.

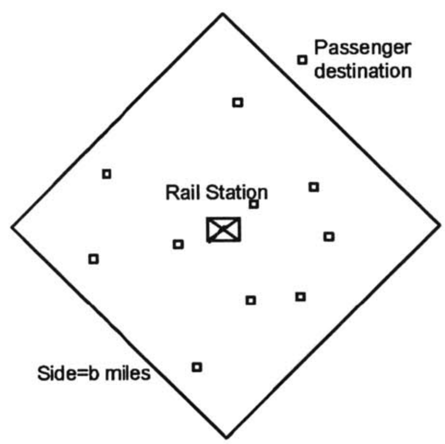


Figure 13: Four-sided diamond service region

In the simulation and numerical comparisons we considered a service region such that $b/v_x = b/v_y = 2.5$ min $= 150$ sec, with a headway of $h = 10$ min $= 600$ sec, and Poisson-distributed customer batch sizes with $\lambda = 20, 40, 60, 80$. Comparisons with the simulation results, when $\lambda = 20$, are shown in Figure 14.

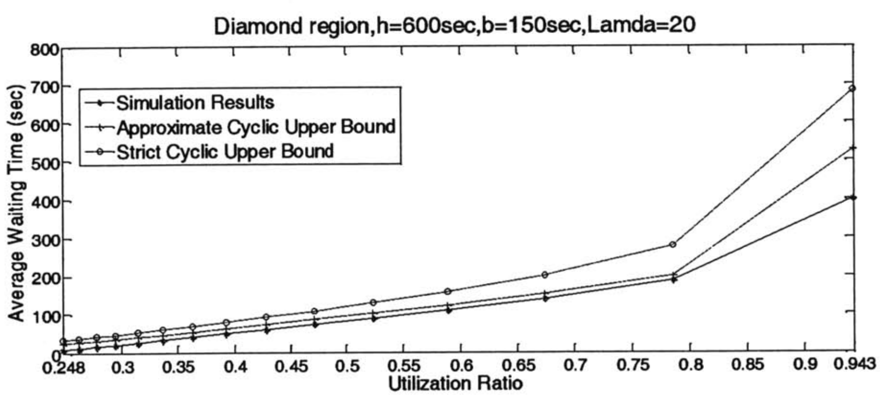


Figure 14: Simulation results and cyclic upper bounds of diamond service region when $b = 150sec, \lambda = 20$

For average waiting times of the order of 1 to 4 minutes, the percent difference between the approximate cyclic upper bound and the simulation results is of the order of 10-20%. The accuracy of the bound is insensitive to the customer demand intensity λ.

2.4.2.T*: Sensitivity analysis of Unit-Capacity, Multi-Vehicle LMP for diamond service region with side *b*

S is individual customer service time (round trip) in the region. We know:

$$f_S(s) = \frac{s}{b^2}, \qquad 0 \le s \le \sqrt{2}b;$$

$$F_S(s) = \frac{s^2}{2b^2}$$

$$E(S) = \frac{2\sqrt{2}}{3}b, E(S^2) = b^2, Var(S) = E(S^2) - E(S)^2 = b^2 - \frac{8}{9}b^2 = \frac{1}{9}b^2, C_S^2 = \frac{1}{8}$$

$$\rho = \frac{E(T)}{h} = \frac{E(N)E(S)}{h} = \frac{2\sqrt{2}b\lambda}{3mh},$$

$$s = E(S)E(N) = \frac{\frac{\sqrt{2}}{3}b\lambda}{m} = \frac{2\sqrt{2}b\lambda}{3m}$$

Strict Lower Bound:

$$W \ge \frac{E(\xi)E(S)E(S^2) + hE(S^2) - 2hE^2(S) - mhE(S^2)}{2E(S)\left(mh - E(\xi)E(S)\right)}$$

$$\ge \frac{\lambda\frac{2\sqrt{2}}{3}bb^2 + hb^2 - 2h(\frac{2\sqrt{2}}{3}b)^2 - mhb^2}{2\frac{2\sqrt{2}}{3}b\left(mh - \lambda\frac{2\sqrt{2}}{3}b\right)} = \frac{b(-h(7+9m) + 6\sqrt{2}b\lambda)}{4(3\sqrt{2}hm - 4b\lambda)}$$

Randomized Upper Bound:

$E(\xi) = \lambda$

$Var(\xi) = \lambda$

$E(\xi^2) = Var(\xi) + E^2(\xi) = \lambda + \lambda^2$

$E(S) = \frac{2\sqrt{2}}{3}b$

$Var(S) = \frac{1}{9}b^2$

$E(S^2) = b^2$

$E(N) = \frac{\lambda}{m}$

$Var(N) = \frac{\lambda}{m}$

$C_a^2 = 0$ (Due to constant batch or macro customer inter-arrival time)

$$C_s^2 = C_t^2 = \frac{E(N)Var(S) + E^2(S)Var(N)}{E^2(N)E^2(S)} = \frac{\frac{\lambda}{m}\frac{1}{9}b^2 + (\frac{2\sqrt{2}}{3}b)^2\frac{\lambda}{m}}{\frac{\lambda^2}{m^2}(\frac{2\sqrt{2}}{3}b)^2} = \frac{9m}{8\lambda}$$

$$s = E(S)E(N) = \frac{\frac{\sqrt{2}}{3}b\lambda}{m} = \frac{2\sqrt{2}b\lambda}{3m}$$

$$\rho = \frac{E(T)}{h} = \frac{E(N)E(S)}{h} = \frac{2\sqrt{2}b\lambda}{3mh}$$

Thus, using the general conclusion, we can obtain a strict upper bound for the average waiting time in the original $D^{\xi}/G/m/\infty$ model:

$$W \le \frac{mhE(S)E(\xi^2) - mhE(S)E(\xi) + mE(S^2)E^2(\xi) - E^2(S)E^3(\xi)}{2m\big(mh - E(\xi)E(S)\big)E(\xi)}$$

$$= \frac{mh\frac{2\sqrt{2}}{3}b(\lambda + \lambda^2) - mh\frac{2\sqrt{2}}{3}b\lambda + mb^2\lambda^2 - (\frac{2\sqrt{2}}{3}b)^2\lambda^3}{2m\left(mh - \lambda\frac{2\sqrt{2}}{3}b\right)\lambda}$$

$$= \frac{b\lambda(9bm + 6\sqrt{2}hm - 8b\lambda)}{6m(3hm - 2\sqrt{2}b\lambda)} \qquad (2.4.2.T.2)$$

Similarly, we have the approximation:

$$W_1 \approx \frac{\rho}{1-\rho} \cdot \frac{(C_a^2 + C_s^2)}{2} s \cdot \exp\left[-\frac{2(1-\rho)(1-C_a^2)^2}{3\rho(C_a^2 + C_s^2)}\right]$$

$$= \frac{\frac{2\sqrt{2}b\lambda}{3mh}}{1 - \frac{2\sqrt{2}b\lambda}{3mh}} \cdot \frac{\frac{9m}{8\lambda}}{2} \frac{2\sqrt{2}b\lambda}{3m} \cdot \exp\left[-\frac{2\left(1 - \frac{2\sqrt{2}b\lambda}{3mh}\right)}{3\frac{2\sqrt{2}b\lambda}{3mh}\frac{9m}{8\lambda}}\right]$$

$$= \frac{3b^2\lambda}{6hm - 4\sqrt{2}b\lambda} \cdot \exp\left[\frac{4}{27}\left(-\frac{3\sqrt{2}h}{b} + \frac{4\lambda}{m}\right)\right]$$

$$W_{D^N/G/1/\infty} = W_1 + \frac{E(S)Var(N) + E(S)E^2(N) - E(S)E(N)}{2E(N)}$$

$$\approx \frac{3b^2\lambda}{6hm - 4\sqrt{2}b\lambda} \cdot \exp\left[\frac{4}{27}\left(-\frac{3\sqrt{2}h}{b} + \frac{4\lambda}{m}\right)\right] + \frac{\frac{2\sqrt{2}}{3}b\frac{\lambda}{m} + \frac{2\sqrt{2}}{3}b\frac{\lambda^2}{m^2} - \frac{2\sqrt{2}}{3}b\frac{\lambda}{m}}{2\frac{\lambda}{m}}$$

$$= \frac{3b^2\lambda}{6hm - 4\sqrt{2}b\lambda} \cdot \exp\left[\frac{4}{27}\left(-\frac{3\sqrt{2}h}{b} + \frac{4\lambda}{m}\right)\right] + \frac{\sqrt{2}b\lambda}{3m}$$

$$W \le W_{D^N/G/1/\infty} \approx \frac{3b^2\lambda}{6hm - 4\sqrt{2}b\lambda} \cdot \exp\left[\frac{4}{27}\left(-\frac{3\sqrt{2}h}{b} + \frac{4\lambda}{m}\right)\right] + \frac{\sqrt{2}b\lambda}{3m} \qquad (2.4.2.T.3)$$

Both the approximate upper bound and the strict upper bound are dimensionally correct. The strict upper bound is larger than the approximate upper bound.

Under heavy traffic,

$$\rho \to 1, 3mh - 2\sqrt{2}b\lambda \to 0, \exp\left[\frac{4}{27}\left(-\frac{3\sqrt{2}h}{b} + \frac{4\lambda}{m}\right)\right] \to 1,$$

the difference between the approximate and strict bounds is reduced to zero.

Cyclic Upper Bound:

$$E(\xi) = \lambda$$

$$Var(\xi) = \lambda$$

$$E(\xi^2) = Var(\xi) + E^2(\xi) = \lambda + \lambda^2$$

$$E(S) = \frac{2\sqrt{2}}{3}b$$

$$Var(S) = \frac{1}{9}b^2$$

$$E(S^2) = b^2$$

$$E(N) = \frac{\lambda}{m}$$

$$Var(N) \leq \frac{4Var(\xi) + m^2}{4m^2} = \frac{4\lambda + m^2}{4m^2}$$

$C_a^2 = 0$ (Due to constant batch or macro customer inter-arrival time)

$$C_s^2 = C_t^2 = \frac{E(N)Var(S) + E^2(S)Var(N)}{E^2(N)E^2(S)} \leq \frac{\frac{\lambda}{m}\frac{1}{9}b^2 + (\frac{2\sqrt{2}}{3}b)^2\frac{4\lambda + m^2}{4m^2}}{\frac{\lambda^2}{m^2}(\frac{2\sqrt{2}}{3}b)^2} = \frac{2m^2 + 8\lambda + m\lambda}{8\lambda^2}$$

$$s = E(S)E(N) = \frac{\frac{\sqrt{2}}{3}b\lambda}{m} = \frac{2\sqrt{2}b\lambda}{3m}$$

$$\rho = \frac{E(T)}{h} = \frac{E(N)E(S)}{h} = \frac{2\sqrt{2}b\lambda}{3mh}$$

Thus, using the general conclusion, we can obtain a strict upper bound for the average waiting time in the original $D^{\xi}/G/m/\infty$ model:

$$W \leq \frac{4mE^2(\xi)E(S^2) - 4E^2(S)E^3(\xi) + 4mhE(S)E(\xi^2) + m^3hE(S) - 4m^2hE(S)E(\xi)}{8m\big(mh - E(\xi)E(S)\big)E(\xi)}$$

$$= \frac{4m\lambda^2 b^2 - 4(\frac{2\sqrt{2}}{3}b)^2\lambda^3 + 4mh\frac{2\sqrt{2}}{3}b(\lambda+\lambda^2) + m^3h\frac{2\sqrt{2}}{3}b - 4m^2h\frac{2\sqrt{2}}{3}b\lambda}{8m\left(mh - \lambda\frac{2\sqrt{2}}{3}b\right)\lambda}$$

$$= \frac{b(2b(9m-8\lambda)\lambda^2 + 3\sqrt{2}hm(m^2 - 4m\lambda + 4\lambda(1+\lambda)))}{12m\lambda(3hm - 2\sqrt{2}b\lambda)} \qquad (2.4.2.T.4)$$

Similarly, we have the approximation:

$$W_1 \approx \frac{\rho}{1-\rho}\cdot\frac{(C_a^2 + C_s^2)}{2}s\cdot\exp\left[-\frac{2(1-\rho)(1-C_a^2)^2}{3\rho(C_a^2+C_s^2)}\right]$$

$$= \frac{\frac{2\sqrt{2}b\lambda}{3mh}}{1-\frac{2\sqrt{2}b\lambda}{3mh}}\frac{\frac{2m^2+8\lambda+m\lambda}{8\lambda^2}}{2}\frac{2\sqrt{2}b\lambda}{3m}\cdot\exp\left[-\frac{2\left(1-\frac{2\sqrt{2}b\lambda}{3mh}\right)}{3\frac{2\sqrt{2}b\lambda}{3mh}\frac{2m^2+8\lambda+m\lambda}{8\lambda^2}}\right]$$

$$= \frac{b^2(2m^2+8\lambda+m\lambda)}{6m(3hm-2\sqrt{2}b\lambda)}\cdot\exp\left[\frac{4\lambda(-3\sqrt{2}hm+4b\lambda)}{3b(2m^2+8\lambda+m\lambda)}\right]$$

$$W_{D^N/G/1/\infty} = W_1 + \frac{E(S)Var(N) + E(S)E^2(N) - E(S)E(N)}{2E(N)}$$

$$\leq \frac{b^2(2m^2+8\lambda+m\lambda)}{6m(3hm-2\sqrt{2}b\lambda)}\cdot\exp\left[\frac{4\lambda(-3\sqrt{2}hm+4b\lambda)}{3b(2m^2+8\lambda+m\lambda)}\right]$$

$$+ \frac{\frac{2\sqrt{2}}{3}b\frac{4\lambda+m^2}{4m^2} + \frac{2\sqrt{2}}{3}b\frac{\lambda^2}{m^2} - \frac{2\sqrt{2}}{3}b\frac{\lambda}{m}}{2\frac{\lambda}{m}}$$

$$= \frac{b^2(2m^2+8\lambda+m\lambda)}{6m(3hm-2\sqrt{2}b\lambda)}\cdot\exp\left[\frac{4\lambda(-3\sqrt{2}hm+4b\lambda)}{3b(2m^2+8\lambda+m\lambda)}\right]$$

$$+ \frac{b(m^2 - 4m\lambda + 4\lambda(1+\lambda))}{6\sqrt{2}m\lambda}$$

$$W \leq W_{D^N/G/1/\infty}$$

$$\leq \frac{b^2(2m^2+8\lambda+m\lambda)}{6m(3hm-2\sqrt{2}b\lambda)}\cdot\exp\left[\frac{4\lambda(-3\sqrt{2}hm+4b\lambda)}{3b(2m^2+8\lambda+m\lambda)}\right]$$

$$+ \frac{b(m^2 - 4m\lambda + 4\lambda(1+\lambda))}{6\sqrt{2}m\lambda} \qquad (2.4.2.T.5)$$

Both the approximate upper bound and the strict upper bound are dimensionally correct. The strict upper bound is larger than the approximate upper bound.

Under heavy traffic,

$$\rho \to 1{,}3mh - 2\sqrt{2}b\lambda \to 0, \exp\left[\frac{4\lambda(-3\sqrt{2}hm + 4b\lambda)}{3b(2m^2 + 8\lambda + m\lambda)}\right] \to 1,$$

the difference between the approximate and strict bounds is reduced to zero.

2.4.3 Rectangular Service Region with Barrier

The service region is next assumed to be rectangular service region that contains an impenetrable barrier to travel. The geometry of the barrier is shown in Figure 15. Technical Section 2.4.3.T contains the theoretical derivations for this case.

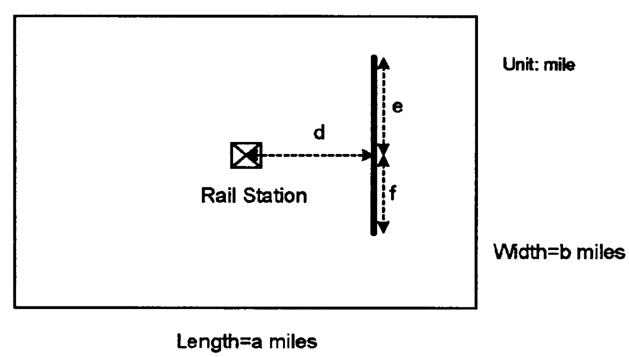


Figure 15: Rectangle service region with barrier inside

In the simulation and numerical comparisons we considered a service region such that $a/v_x = 2.5\ min = 150\ sec, b/v_y = 2$ min $= 120$ sec , $d/v_x = 0.625$ min $= 37.5$ sec , $e/v_y = 0.5$ min $= 30$ sec , $f/v_y = 0.25$ min $= 15$ sec, with headway of $h = 10$ min $= 600$ sec, and Poisson-distributed passenger batch sizes of $\lambda = 20, 40, 60, 80$. The simulation results when $\lambda = 20$ are shown in Figures 16.

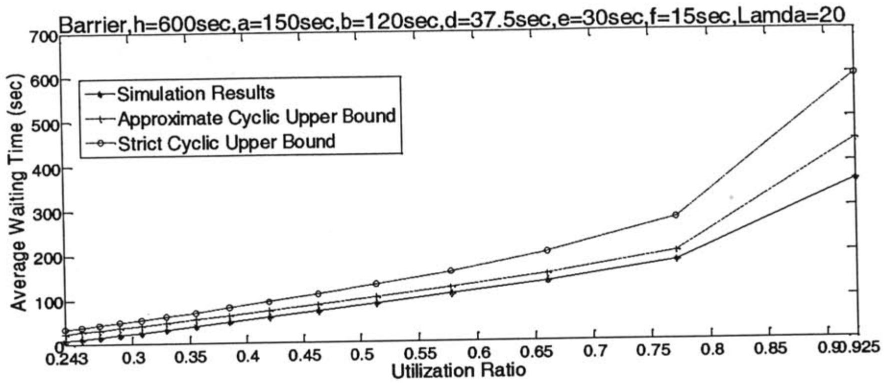


Figure 16: Simulation results and cyclic upper bounds, rectangle service region with barrier, when $\lambda = 20$.

For average waiting times of the order of 1 to 4 minutes, the percent difference between the approximate cyclic upper bound and the simulation results is again of the order of 10-20%, and the accuracy of the bound was insensitive to the customer demand intensity λ.

Overall, the sensitivity analysis of this section, suggests that the strict cyclic upper bound and the approximate cyclic upper bound remain valid and provide good estimates of performance for a wide range of customer demand rates and for differently shaped compact and convex service regions.

2.4.3.T*: Sensitivity analysis of Unit-Capacity, Multi-Vehicle LMP for rectangle service region with barrier

S is individual customer service time (round trip) in the region. We divide the region to six different areas, then compute $E(S), E(S^2), Var(S)$, using total expectation law. The area division is illustrated as follows:

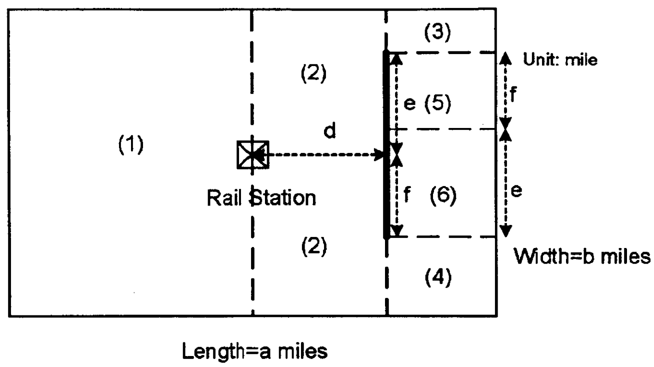


Figure 17: Area division for the rectangle with barrier

$$E(S) = \sum_{i=1}^{6} P(Area\ i)E(S|Area\ i)$$

$$= \frac{\frac{a}{2}b}{ab}2\left(\frac{a}{4}+\frac{b}{4}\right) + \frac{bd}{ab}2\left(\frac{d}{2}+\frac{b}{4}\right) + \frac{(\frac{b}{2}-e)(\frac{a}{2}-d)}{ab}2\left(d+e+\frac{\frac{a}{2}-d}{2}+\frac{\frac{b}{2}-e}{2}\right)$$

$$+\frac{(\frac{b}{2}-f)(\frac{a}{2}-d)}{ab}2\left(d+f+\frac{\frac{a}{2}-d}{2}+\frac{\frac{b}{2}-f}{2}\right)$$

$$+\frac{f(\frac{a}{2}-d)}{ab}2\left(d+e+\frac{\frac{a}{2}-d}{2}+\frac{f}{2}\right)+\frac{e(\frac{a}{2}-d)}{ab}2\left(d+f+\frac{\frac{a}{2}-d}{2}+\frac{e}{2}\right)$$

$$= \frac{1}{2}\left(a+b+\frac{4ef}{b}-\frac{8def}{ab}\right) \qquad (2.4.3.T.1)$$

$$E(S|corner\ to\ a \times b\ Rectangle, round\ trip) = a + b$$

$$E(S^2|corner\ to\ a \times b\ Rectangle, round\ trip) = \frac{4a^2}{3} + 2ab + \frac{4b^2}{3}$$

$$E(S|Point\ g\ away\ from\ corner\ to\ a \times b\ Rectangle, round\ trip) = 2g + a + b$$

$$E(S^2|Point\ g\ away\ from\ corner\ to\ a \times b\ Rectangle, round\ trip) = E((2g + S_e)^2)$$

$$= E\left(4g^2 + {S_e}^2 + 4gS_e\right) = 4g^2 + 4gE(S_e) + E\left({S_e}^2\right)$$

$$= 4g^2 + 4g(a+b) + \frac{4a^2}{3} + 2ab + \frac{4b^2}{3}$$

$$= 4g^2 + 4g(a+b) + \frac{4a^2}{3} + 2ab + \frac{4b^2}{3}$$

$$E(S^2|Area\ 1) = \frac{a^2}{3} + \frac{ab}{2} + \frac{b^2}{3}$$

$$E(S^2|Area\ 2) = \frac{b^2}{3} + bd + \frac{4d^2}{3}$$

$$E(S^2|Area\ 3) = 4g^2 + 4g(a+b) + \frac{4a^2}{3} + 2ab + \frac{4b^2}{3}$$

$$= 4(d+e)^2 + 4(d+e)\left(\left(\frac{a}{2} - d\right) + \left(\frac{b}{2} - e\right)\right) + \frac{4(\frac{a}{2} - d)^2}{3}$$

$$+ 2\left(\frac{a}{2} - d\right)\left(\frac{b}{2} - e\right) + \frac{4\left(\frac{b}{2} - e\right)^2}{3}$$

$$= \frac{1}{6}(2a^2 + a(3b + 4d + 6e) + 2(b^2 + 3bd + 4d^2 + 2be + 6de + 4e^2))$$

$$E(S^2|Area\ 4) = 4g^2 + 4g(a+b) + \frac{4a^2}{3} + 2ab + \frac{4b^2}{3}$$

$$= 4(d+f)^2 + 4(d+f)\left(\left(\frac{a}{2} - d\right) + \left(\frac{b}{2} - f\right)\right) + \frac{4(\frac{a}{2} - d)^2}{3}$$

$$+ 2\left(\frac{a}{2} - d\right)\left(\frac{b}{2} - f\right) + \frac{4\left(\frac{b}{2} - f\right)^2}{3}$$

$$= \frac{1}{6}(2a^2 + a(3b + 4d + 6f) + 2(b^2 + 3bd + 4d^2 + 2bf + 6df + 4f^2))$$

$$E(S^2|Area\ 5) = 4g^2 + 4g(a+b) + \frac{4a^2}{3} + 2ab + \frac{4b^2}{3}$$

$$= 4(d+e)^2 + 4(d+e)\left(\left(\frac{a}{2} - d\right) + f\right) + \frac{4(\frac{a}{2} - d)^2}{3} + 2\left(\frac{a}{2} - d\right)f + \frac{4f^2}{3}$$

$$= \frac{1}{3}(a^2 + a(2d + 6e + 3f) + 2(2d^2 + 6de + 6e^2 + 3df + 6ef + 2f^2))$$

$$E(S^2|Area\ 6) = 4g^2 + 4g(a+b) + \frac{4a^2}{3} + 2ab + \frac{4b^2}{3}$$

$$= 4(d+f)^2 + 4(d+f)\left(\left(\frac{a}{2} - d\right) + e\right) + \frac{4(\frac{a}{2} - d)^2}{3} + 2\left(\frac{a}{2} - d\right)e + \frac{4e^2}{3}$$

$$= \frac{1}{3}(a^2 + a(2d + 3e + 6f) + 2(2d^2 + 3de + 2e^2 + 6df + 6ef + 6f^2))$$

Therefore,

$$E(S^2) = \sum_{i=1}^{6} P(Area\ i)E(S^2|Area\ i)$$

$$= \frac{\frac{a}{2}b}{ab}(\frac{a^2}{3}+\frac{ab}{2}+\frac{b^2}{3}) + \frac{bd}{ab}(\frac{b^2}{3}+bd+\frac{4d^2}{3}) + \frac{\left(\frac{b}{2}-e\right)\left(\frac{a}{2}-d\right)}{ab}(\frac{1}{6}(2a^2$$
$$+ a(3b+4d+6e) + 2(b^2+3bd+4d^2+2be+6de+4e^2)))$$
$$+ \frac{\left(\frac{b}{2}-f\right)\left(\frac{a}{2}-d\right)}{ab}(\frac{1}{6}(2a^2+a(3b+4d+6f)+2(b^2+3bd+4d^2+2bf$$
$$+ 6df + 4f^2))) + \frac{f\left(\frac{a}{2}-d\right)}{ab}(\frac{1}{3}(a^2+a(2d+6e+3f)+2(2d^2+6de+6e^2$$
$$+ 3df + 6ef + 2f^2))) + \frac{e\left(\frac{a}{2}-d\right)}{ab}(\frac{1}{3}(a^2+a(2d+3e+6f)+2(2d^2+3de$$
$$+ 2e^2 + 6df + 6ef + 6f^2)))$$
$$= \frac{2a^3b - 48def(d+e+f) + 3a^2(b^2+4ef) + 2a(b^3+12ef(e+f))}{6ab}$$

(2.4.3.T.2)

$$Var(S) = E(S^2) - E^2(S)$$
$$= \frac{2a^3b - 48def(d+e+f) + 3a^2(b^2+4ef) + 2a(b^3+12ef(e+f))}{6ab}$$
$$- (\frac{1}{2}\left(a+b+\frac{4ef}{b}-\frac{8def}{ab}\right))^2 \quad (2.4.3.T.3)$$

$$C_s^2 = \frac{Var(S)}{E^2(S)}$$
$$= \frac{\frac{2a^3b - 48def(d+e+f) + 3a^2(b^2+4ef) + 2a(b^3+12ef(e+f))}{6ab} - (\frac{1}{2}\left(a+b+\frac{4ef}{b}-\frac{8def}{ab}\right))^2}{(\frac{1}{2}\left(a+b+\frac{4ef}{b}-\frac{8def}{ab}\right))^2}$$

We consider a specific example with the following geometry:

$$a = \frac{5}{4}b, d = \frac{a}{4} = \frac{5}{16}b, e = \frac{b}{4}, f = \frac{b}{8}$$

S is individual customer service time (round trip) in the region. From the conclusion above, we obtain:

$$E(S) = \frac{1}{2}\left(a + b + \frac{4ef}{b} - \frac{8def}{ab}\right) = \frac{37b}{32}$$

$$Var(S) = \frac{2a^3b - 48def(d+e+f) + 3a^2(b^2+4ef) + 2a(b^3 + 12ef(e+f))}{6ab} - (\frac{1}{2}\left(a + b + \frac{4ef}{b} - \frac{8def}{ab}\right))^2 = \frac{689b^2}{3072},$$

$$E(S^2) = \frac{2a^3b - 48def(d+e+f) + 3a^2(b^2+4ef) + 2a(b^3 + 12ef(e+f))}{6ab} = \frac{1199b^2}{768},$$

$$C_s^2 = \frac{\frac{689b^2}{3072}}{(\frac{37b}{32})^2} = \frac{689}{4107},$$

$$\rho = \frac{E(T)}{h} = \frac{E(N)\frac{37b}{32}}{h} = \frac{37b\lambda}{32mh}$$

$$s = E(S)E(N) = \frac{37b\lambda}{32m}$$

Strict Lower Bound:

$$W \geq \frac{E(\xi)E(S)E(S^2) + hE(S^2) - 2hE^2(S) - mhE(S^2)}{2E(S)\big(mh - E(\xi)E(S)\big)}$$

$$\geq \frac{\lambda\frac{37b}{32}\frac{1199b^2}{768} + h\frac{1199b^2}{768} - 2h(\frac{37b}{32})^2 - mh\frac{1199b^2}{768}}{2\frac{37b}{32}\left(mh - \lambda\frac{37b}{32}\right)}$$

$$= \frac{b(-16h(1709 + 2398m) + 44363b\lambda)}{1776(32hm - 37b\lambda)} \qquad (2.4.3.T.4)$$

Randomized Upper Bound:

$$E(\xi) = \lambda$$

$$Var(\xi) = \lambda$$

$$E(\xi^2) = Var(\xi) + E^2(\xi) = \lambda + \lambda^2$$

$$E(S) = \frac{37b}{32}$$

$$Var(S) = \frac{689b^2}{3072},$$

$$E(S^2) = \frac{1199b^2}{768},$$

$$E(N) = \frac{\lambda}{m}$$

$$Var(N) = \frac{\lambda}{m}$$

$C_a^2 = 0$ (Due to constant batch or macro customer inter-arrival time)

$$C_s^2 = C_t^2 = \frac{E(N)Var(S) + E^2(S)Var(N)}{E^2(N)E^2(S)} = \frac{\frac{\lambda}{m}\frac{689b^2}{3072} + (\frac{37b}{32})^2\frac{\lambda}{m}}{\frac{\lambda^2}{m^2}(\frac{37b}{32})^2} = \frac{4796m}{4107\lambda}$$

$$s = E(S)E(N) = \frac{\frac{37b}{32}\lambda}{m} = \frac{37b\lambda}{32m}$$

$$\rho = \frac{E(T)}{h} = \frac{E(N)\frac{37b}{32}}{h} = \frac{37b\lambda}{32mh}$$

Thus, using the general conclusion, we can obtain a strict upper bound for the average waiting time in the original $D^{\xi}/G/m/\infty$ model:

$$W \le \frac{mhE(S)E(\xi^2) - mhE(S)E(\xi) + mE(S^2)E^2(\xi) - E^2(S)E^3(\xi)}{2m\left(mh - E(\xi)E(S)\right)E(\xi)}$$

$$= \frac{mh\frac{37b}{32}(\lambda + \lambda^2) - mh\frac{37b}{32}\lambda + m\frac{1199b^2}{768}\lambda^2 - (\frac{37b}{32})^2\lambda^3}{2m\left(mh - \lambda\frac{37b}{32}\right)\lambda}$$

$$= \frac{b\lambda(4796bm + 3552hm - 4107b\lambda)}{192m(32hm - 37b\lambda)} \qquad (2.4.3.T.5)$$

Similarly, we have the approximation:

$$W_1 \approx \frac{\rho}{1-\rho} \cdot \frac{(C_a^2 + C_s^2)}{2} s \cdot \exp\left[-\frac{2(1-\rho)(1-C_a^2)^2}{3\rho(C_a^2 + C_s^2)}\right]$$

$$= \frac{\frac{37b\lambda}{32mh}}{1-\frac{37b\lambda}{32mh}} \cdot \frac{\frac{4796m}{4107\lambda}}{2} \frac{37b\lambda}{32m} \cdot \exp\left[-\frac{2\left(1-\frac{37b\lambda}{32mh}\right)}{3\frac{37b\lambda}{32mh}\frac{4796m}{4107\lambda}}\right]$$

$$= \frac{1199b^2\lambda}{48(32hm - 37b\lambda)} \cdot \exp\left[\frac{37(-32hm + 37b\lambda)}{2398bm}\right]$$

$$W_{D^N/G/1/\infty} = W_1 + \frac{E(S)Var(N) + E(S)E^2(N) - E(S)E(N)}{2E(N)}$$

$$\approx \frac{1199b^2\lambda}{48(32hm - 37b\lambda)} \cdot \exp\left[\frac{37(-32hm + 37b\lambda)}{2398bm}\right] + \frac{\frac{37b}{32}\frac{\lambda}{m} + \frac{37b}{32}\frac{\lambda^2}{m^2} - \frac{37b}{32}\frac{\lambda}{m}}{2\frac{\lambda}{m}}$$

$$= \frac{1199b^2\lambda}{48(32hm - 37b\lambda)} \cdot \exp\left[\frac{37(-32hm + 37b\lambda)}{2398bm}\right] + \frac{37b\lambda}{64m}$$

$$W \le W_{D^N/G/1/\infty} \approx \frac{1199b^2\lambda}{48(32hm - 37b\lambda)} \cdot \exp\left[\frac{37(-32hm + 37b\lambda)}{2398bm}\right] + \frac{37b\lambda}{64m} \qquad (2.4.3.T.6)$$

Both the approximate upper bound and the strict upper bound are dimensionally correct. The strict upper bound is larger than the approximate upper bound.

Under heavy traffic,

$$\rho \to 1, 32mh - 37b\lambda \to 0, \exp\left[\frac{37(-32hm + 37b\lambda)}{2398bm}\right] \to 1,$$

the difference between the approximate and strict bounds is reduced to zero.

Cyclic Upper Bound:

$E(\xi) = \lambda$

$Var(\xi) = \lambda$

$E(\xi^2) = Var(\xi) + E^2(\xi) = \lambda + \lambda^2$

$E(S) = \frac{37b}{32}$

$$Var(S) = \frac{689b^2}{3072},$$

$$E(S^2) = \frac{1199b^2}{768},$$

$$E(N) = \frac{\lambda}{m}$$

$$Var(N) \le \frac{4Var(\xi) + m^2}{4m^2} = \frac{4\lambda + m^2}{4m^2}$$

$C_a^2 = 0$ (Due to constant batch or macro customer inter-arrival time)

$$C_s^2 = C_t^2 = \frac{E(N)Var(S) + E^2(S)Var(N)}{E^2(N)E^2(S)} \le \frac{\frac{\lambda}{m}\frac{689b^2}{3072} + (\frac{37b}{32})^2\frac{4\lambda + m^2}{4m^2}}{\frac{\lambda^2}{m^2}(\frac{37b}{32})^2} = \frac{m^2}{4\lambda^2} + \frac{1}{\lambda} + \frac{689m}{4107\lambda}$$

$$s = E(S)E(N) = \frac{\frac{37b}{32}\lambda}{m} = \frac{37b\lambda}{32m}$$

$$\rho = \frac{E(T)}{h} = \frac{E(N)\frac{37b}{32}}{h} = \frac{37b\lambda}{32mh}$$

Thus, using the general conclusion, we can obtain a strict upper bound for the average waiting time in the original $D^\xi/G/m/\infty$ model:

$$W \le \frac{4mE^2(\xi)E(S^2) - 4E^2(S)E^3(\xi) + 4mhE(S)E(\xi^2) + m^3hE(S) - 4m^2hE(S)E(\xi)}{8m(mh - E(\xi)E(S))E(\xi)}$$

$$= \frac{4m\lambda^2\frac{1199b^2}{768} - 4(\frac{37b}{32})^2\lambda^3 + 4mh\frac{37b}{32}(\lambda + \lambda^2) + m^3h\frac{37b}{32} - 4m^2h\frac{37b}{32}\lambda}{8m\left(mh - \lambda\frac{37b}{32}\right)\lambda}$$

$$= \frac{b(b(4796m - 4107\lambda)\lambda^2 + 888hm(m^2 - 4m\lambda + 4\lambda(1+\lambda)))}{192m\lambda(32hm - 37b\lambda)} \quad (2.4.3.T.7)$$

Similarly, we have the approximation:

$$W_1 \approx \frac{\rho}{1-\rho}\cdot\frac{(C_a^2+C_s^2)}{2}s\cdot\exp\left[-\frac{2(1-\rho)(1-C_a^2)^2}{3\rho(C_a^2+C_s^2)}\right]$$

$$=\frac{\frac{37b\lambda}{32mh}}{1-\frac{37b\lambda}{32mh}}\frac{\frac{m^2}{4\lambda^2}+\frac{1}{\lambda}+\frac{689m}{4107\lambda}}{2}\frac{37b\lambda}{32m}\cdot\exp\left[-\frac{2\left(1-\frac{37b\lambda}{32mh}\right)}{3\frac{37b\lambda}{32mh}\left(\frac{m^2}{4\lambda^2}+\frac{1}{\lambda}+\frac{689m}{4107\lambda}\right)}\right]$$

$$=\frac{b^2(4107m^2+16428\lambda+2756m\lambda)}{768m(32hm-37b\lambda)}$$

$$\cdot\exp\left[\frac{296\lambda(-32hm+37b\lambda)}{b(4107m^2+16428\lambda+2756m\lambda)}\right]$$

$$W_{D^N/G/1/\infty}=W_1+\frac{E(S)Var(N)+E(S)E^2(N)-E(S)E(N)}{2E(N)}$$

$$\leq\frac{b^2(4107m^2+16428\lambda+2756m\lambda)}{768m(32hm-37b\lambda)}$$

$$\cdot\exp\left[\frac{296\lambda(-32hm+37b\lambda)}{b(4107m^2+16428\lambda+2756m\lambda)}\right]+\frac{\frac{37b}{32}\frac{4\lambda+m^2}{4m^2}+\frac{37b}{32}\frac{\lambda^2}{m^2}-\frac{37b}{32}\frac{\lambda}{m}}{2\frac{\lambda}{m}}$$

$$=\frac{b^2(4107m^2+16428\lambda+2756m\lambda)}{768m(32hm-37b\lambda)}$$

$$\cdot\exp\left[\frac{296\lambda(-32hm+37b\lambda)}{b(4107m^2+16428\lambda+2756m\lambda)}\right]+\frac{37b(m^2-4m\lambda+4\lambda(1+\lambda))}{256m\lambda}$$

$$W\leq W_{D^N/G/1/\infty}$$

$$\leq\frac{b^2(4107m^2+16428\lambda+2756m\lambda)}{768m(32hm-37b\lambda)}$$

$$\cdot\exp\left[\frac{296\lambda(-32hm+37b\lambda)}{b(4107m^2+16428\lambda+2756m\lambda)}\right]+\frac{37b(m^2-4m\lambda+4\lambda(1+\lambda))}{256m\lambda}$$

Both the approximate upper bound and the strict upper bound are dimensionally correct. The strict upper bound is larger than the approximate upper bound.

Under heavy traffic,

$$\rho\rightarrow 1, 32mh-37b\lambda\rightarrow 0, \exp\left[\frac{296\lambda(-32hm+37b\lambda)}{b(4107m^2+16428\lambda+2756m\lambda)}\right]\rightarrow 1$$

the difference between the approximate and strict bounds is reduced to zero.

3. General-Capacity, Multi-Vehicle LMP: Upper Bounds and Approximations

In this Chapter we consider the General-Capacity, Multi-Vehicle LMP, in which both the vehicle capacity, c, and the number of vehicles, m, are arbitrary positive integers. The vehicles will now travel along more complicated routes than in the $c = 1$ case to deliver customers to their destinations. In practice, one would expect the vehicle capacity to be a small number of the order of 4 to 10 customers – unless the LMTS fleet consists of bus-size vehicles, in which case the methodologies laid out in this thesis are less applicable.

As explained in Chapter 1, the General-Capacity, Multi-Vehicle LMTS will be viewed as a spatially distributed queuing system in which the service times are equal to the amount of time it takes to complete a customer delivery tour and return to the train station – see also Figure 3. Vehicle routing and path choice issues must therefore be addressed in this connection. This is done in this Chapter, which also summarizes the bounds and approximations we have obtained.

The approach to be described consists of the following three steps: (i) customers are partitioned into clusters with the size of each cluster no larger than the vehicle capacity, c; (ii) each cluster is assigned to a vehicle and a delivery route is designed for each vehicle; (iii) using the service times (i.e., tour durations) computed in the previous step, the (appropriately modified) queuing results from the Unit-Capacity model of Chapter 2 are then applied to estimate system performance. The performance measures we shall concentrate on include average waiting time until boarding a vehicle and average time until delivery to destination, i.e., the sum of the time spent waiting to board a vehicle and of the time spent riding until delivery.

3.1 Approximating the Expectation and Variance of Tour Lengths

Since we are looking for widely applicable approximations and bounds on system performance and not for exact expressions, we have selected a “greedy” partitioning strategy for assigning customers to vehicles. Specifically, we partition customers in each arriving batch simply according to their order of arrival at the station. In other words, Vehicle 1 serves customers $1, 2, \ldots, c$ in a single tour, Vehicle 2 serves customers $c + 1$, $c + 2, \ldots, 2c$ in a single tour, and so on. If we consider the c customers served by one vehicle as a single request for service, the number of service requests after the arrival of each train is given by ξ/c, when the size of an arriving batch is ξ.

For the routing step, we also use a "greedy routing strategy" – which, however, is refined subsequently, in the manner described later in this Chapter. Upon leaving the rail station with c customers on board, the vehicle will first deliver the customer whose destination is closest to the station, denoted as Point A in Figure 18, then the customer whose destination is closest to point A (i.e., Point B in Figure 18) and so forth. Finally, after delivering the last customer (Point F) the vehicle will return to the rail station. Thus, we construct a vehicle tour using essentially a "Nearest Neighbor" (NN) heuristic approach. The reason for following this sub-optimal routing strategy is that it is mathematically feasible to compute approximately both the expected length and the variance of the length of a NN tour that delivers c customers and returns to the rail station. Both of these quantities (expected length and variance of the length) are necessary if one is to apply the queuing expressions derived in Chapter 2.

A better alternative would have been to find the Hamiltonian tour, i.e., the optimal "Traveling Salesman" tour (TST), through the $c + 1$ points (customer destinations plus rail station) to be visited. However, we are not aware of any simple explicit expressions for the variance of the length of TST tours. We have therefore opted for the NN-based routing approach. We have, however, attempted to correct the expressions for "expected length" and "variance of length" derived through the NN-based approach, by comparing these with corresponding estimates (expectation and variance) obtained through many numerical experiments.

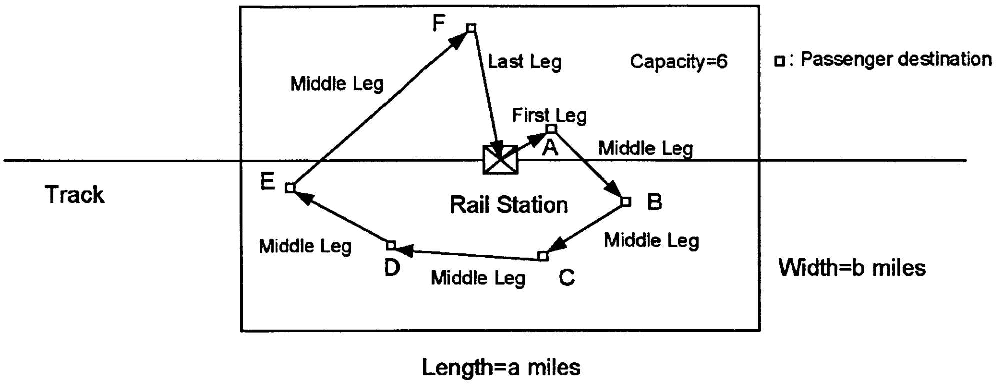


Figure 18: Greedy routing strategy for the General-Capacity, Multi-Vehicle LMP

The tour shown in Figure 18 consists of one *First Leg*, $c - 1$ *Middle Legs*, and one *Last Leg*. The expected length of the entire route is then given by

$$E(S_E) = s_{First\ Leg} + s_{middle,c-1} + \cdots + s_{middle,1} + s_{Last\ Leg} \quad (11)$$

where the notation $s_{First\ Leg}$ and $s_{Last\ Leg}$ denotes, respectively, the expected length of the first and last legs of the tour, while $s_{middle,k}$ denotes the expected distance between the destination of the last customer delivered and the nearest destination of k remaining customers still to be delivered. For example, $s_{middle,c-1}$ denotes the distance between the first of the customers delivered (i.e., the nearest one to the rail station) and the nearest destination among the destinations of the remaining $c - 1$ customers still to be delivered.

The variance of the length of the entire service route can be similarly approximated as

$$VAR_E = VAR_{First\ Leg} + VAR_{middle,c-1} + \cdots + VAR_{middle,1} + VAR_{Last\ Leg}, \quad (12)$$

where VAR denotes a variance and the subscripts can be interpreted in exactly the same way as the subscripts of the expectations, s, above. Finally, the second moment of the length of the entire service tour is given by $SQ_E = E(S_E^2) = VAR_E + (E(S_E))^2$.

The above estimates of the moments and variance of the service tour can be converted into time units, if one is given information about the speed of travel in the region of interest. To simplify this conversion, we shall continue to assume here that travel speed is constant and equal to 1 throughout the region.

We have derived approximate expressions for $E(S_E)$, VAR_E , and SQ_E assuming a right-angle travel metric and a rectangular service region of size $a \times b$. With the NN ("greedy") routing strategy, the length of the first leg of the delivery tour is the distance from the rail station to the nearest of c random points (c random customer destinations), while the last leg is the distance from another (approximately) random point (the destination of the final customer served in the tour) back to the rail station. It is not difficult to derive the expectation and variance of these distances as shown in Technical Section 3.1.T1 and 3.1.T2, respectively.

3.1.T1*: Expectation and variance for first leg under greedy routing strategy

First Leg: Service time from the rail station to the nearest of c random points (c random customer destinations).

In the X axis, random point is uniformly distributed from 0 to $a/2$:

$$f_X(x) = \frac{2}{a}, x \in [0, a/2]$$

In the Y axis, random point is uniformly distributed from 0 to $b/2$:

$$f_Y(y) = \frac{2}{b}, y \in [0, b/2]$$

$$S = X + Y$$

$$f_S(s) = \begin{cases} \frac{4}{ab}s, & s \in \left[0, \frac{b}{2}\right] \\ \frac{2}{a}, & s \in \left[\frac{b}{2}, \frac{a}{2}\right] \\ \frac{2}{a} + \frac{2}{b} - \frac{4}{ab}s, & s \in \left[\frac{a}{2}, \frac{a+b}{2}\right] \end{cases} \quad (3.1.T1.1)$$

$$F_S(s) = \int_0^s f_S(x)dx = \begin{cases} \frac{2}{ab}s^2, & x \in \left[0, \frac{b}{2}\right] \\ \frac{4s-b}{2a}, & x \in \left[\frac{b}{2}, \frac{a}{2}\right] \\ \frac{-a^2 - b^2 - 4s^2 + 4as + 4bs}{2ab}, & x \in \left[\frac{a}{2}, \frac{a+b}{2}\right] \end{cases} \quad (3.1.T1.2)$$

$S_F = \min(S_1, S_2, \ldots, S_c)$, where S_i is identically distributed S.

$$F_{S_F}(s) = 1 - P(S_F > s) = 1 - P(\min(S_1, S_2, \ldots, S_c) > s) = 1 - P(S_1 > s) \ldots P(S_c > s)$$

$$= 1 - (1 - F_S(s))^c$$

$$f_{S_F}(s) = \frac{dF_{S_F}(s)}{ds} = cf_S(s)(1 - F_S(s))^{c-1}$$

Therefore, the expectation of $S_F(a > b)$:

$$s_{First\ Leg} = E(S_F) = \int_0^{\infty} s f_{S_F}(s) ds = \int_0^{\infty} s c f_S(s)(1 - F_S(s))^{c-1} ds$$

$$= \frac{1}{4}\left(\frac{2^{-c}a^{-c}((2a-b)^{1+c} - b^{1+c})}{1+c} + \frac{2^{1-c}b(\frac{b}{a})^c}{1+2c}\right.$$

$$\left. + 2b\text{Hypergeometric2F1}[\frac{1}{2}, -c, \frac{3}{2}, \frac{b}{2a}]\right) \qquad (3.1.T1.3)$$

where Hypergeometric2F1 $= F_1(a, b; c; z) = \sum_{k=0}^{\infty}(a)_k (b)_k/(c)_k z^k/k!$, $(a)_k = a(a+1)\ldots(a+k-1) = \Gamma(a+k)/\Gamma(a)$, $(b)_k = b(b+1)\ldots(b+k-1) = \Gamma(b+k)/\Gamma(b)$, $(c)_k = c(c+1)\ldots(c+k-1) = \Gamma(c+k)/\Gamma(c)$.

The second moment of $S_F(a > b)$:

$$SQ_{First\ Leg} = E(S_F{}^2) = \int_0^{\infty} s^2 f_{S_F}(s) ds = \int_0^{\infty} s^2 c f_S(s)\left(1 - F_S(s)\right)^{c-1} ds$$

$$= 2^{-3-c}\left(\frac{2b(2^{1+c}a - (2 - \frac{b}{a})^c(2a + bc))}{1+c}\right.$$

$$+ \frac{2(\frac{b}{a})^c(b^2 + 2ab(1+c) + a^2(1 + 3c + 2c^2))}{1 + 3c + 2c^2}$$

$$+ \frac{1}{(1+c)(2+c)}a^{-c}((2a-b)^c((2a+b)^2 + 4b(a+b)c + 2b^2c^2) - b^c(b^2$$

$$\left. + 2ab(2+c) + 2a^2(1+c)(2+c))\right)) \qquad (3.1.T1.4)$$

$$Var(S_F) = E(S_F{}^2) - E(S_F)^2$$

$$= 2^{-3-c}(\frac{2b(2^{1+c}a - (2-\frac{b}{a})^c(2a+bc))}{1+c}$$

$$+\frac{2(\frac{b}{a})^c(b^2+2ab(1+c)+a^2(1+3c+2c^2))}{1+3c+2c^2}$$

$$+\frac{1}{(1+c)(2+c)}a^{-c}((2a-b)^c((2a+b)^2+4b(a+b)c+2b^2c^2)-b^c(b^2$$

$$+2ab(2+c)+2a^2(1+c)(2+c)))$$

$$-2^{-4-2c}(\frac{2\left(\frac{b}{a}\right)^c(a+b+2ac)}{1+2c}$$

$$+\frac{a^{-c}\left((2a-b)^c(2a+b+2bc)-b^c(b+2a(1+c))\right)}{1+c}$$

$$+\frac{1}{-1+4c^2}4\left(2-\frac{b}{a}\right)^c c(-2a+b-2bc$$

$$+2^{1+c}a(\frac{a}{2a-b})^c\text{Hypergeometric2F1}[-\frac{1}{2},1-c,\frac{1}{2},\frac{b}{2a}]))^2 \qquad (3.1.T1.5)$$

When the region is square, i.e., $a = b$,

$$f_S(s) = \begin{cases} \frac{4}{b^2}s, & s \in \left[0,\frac{b}{2}\right] \\ \frac{4}{b}-\frac{4}{b^2}s, & s \in \left[\frac{b}{2},b\right] \end{cases}$$

$$F_S(s) = \int_0^s f_S(x)dx = \begin{cases} \frac{2s^2}{b^2}, & x \in \left[0,\frac{b}{2}\right] \\ \frac{-b^2+4bs-2s^2}{b^2}, & x \in \left[\frac{b}{2},b\right] \end{cases}$$

$S_F = \min(S_1,S_2,\ldots,S_c)$, where S_i is identically distributed S.

$$F_{S_F}(s) = 1 - P(S_F > s) = 1 - P(\min(S_1,S_2,\ldots,S_c) > s) = 1 - P(S_1 > s)\ldots P(S_c > s)$$

$$= 1 - (1 - F_S(s))^c$$

$$f_{S_F}(s) = \frac{dF_{S_F}(s)}{ds} = cf_S(s)(1 - F_S(s))^{c-1}$$

Therefore, the expectation of $S_F(a = b)$:

$$s_{First\ Leg} = E(S_F) = \int_0^\infty s f_{S_F}(s) ds = \int_0^\infty s c f_S(s)(1 - F_S(s))^{c-1} ds$$

$$= \frac{2^{-2-c} b(1 + (1 + 2c)\text{Hypergeometric2F1}[1, -1 - c, \frac{1}{2}, -1])}{1 + 3c + 2c^2} \quad (3.1.\text{T}1.6)$$

The second moment of $S_F(a = b)$:

$$SQ_{First\ Leg} = E(S_F{}^2) = \int_0^\infty s^2 f_{S_F}(s) ds = \int_0^\infty s^2 c f_S(s)\left(1 - F_S(s)\right)^{c-1} ds$$

$$= \frac{2^{-1-c} b^2 (1 + 2^c + 2^{1+c} c)}{1 + 3c + 2c^2} \quad (3.1.\text{T}1.7)$$

The variance of $S_F(a = b)$:

$$Var(S_F) = E(S_F{}^2) - E(S_F)^2$$

$$= \frac{2^{-1-c} b^2 (1 + 2^c + 2^{1+c} c)}{1 + 3c + 2c^2}$$

$$- \frac{2^{-4-2c} b^2 (1 + (1 + 2c)\text{Hypergeometric2F1}[1, -1 - c, \frac{1}{2}, -1])^2}{(1 + 3c + 2c^2)^2} \quad (3.1.\text{T}1.8)$$

3.1.T2*: Expectation and variance for last leg under greedy routing strategy

Last Leg: Service time from one (approximately) random point (the destination of the final customer served in the tour) back to the rail station.

In the X axis, random point is uniformly distributed from 0 to $a/2$:

$$f_X(x) = \frac{2}{a}, x \in [0, a/2]$$

In the Y axis, random point is uniformly distributed from 0 to $b/2$:

$$f_Y(y) = \frac{2}{b}, y \in [0, b/2]$$

$$S = X + Y$$

$$f_S(s) = \begin{cases} \frac{4}{ab}s, & x \in \left[0, \frac{b}{2}\right] \\ \frac{2}{a}, & x \in \left[\frac{b}{2}, \frac{a}{2}\right] \\ \frac{2}{a} + \frac{2}{b} - \frac{4}{ab}s, & x \in \left[\frac{a}{2}, \frac{a+b}{2}\right] \end{cases}$$

$$s_{Last\ Leg} = E(S) = \frac{a+b}{4}, \tag{3.1.T2.1}$$

$$VAR_{Last\ Leg} = Var(S) = \frac{a^2+b^2}{48} \tag{3.1.T2.2}$$

$$SQ_{Last\ Leg} = E(S^2) = \frac{2a^2+2b^2+3ab}{24} \tag{3.1.T2.3}$$

When the region is square, i.e., $a = b$,

$$s_{Last\ Leg} = \frac{b}{2}, VAR_{Last\ Leg} = \frac{b^2}{24}, SQ_{Last\ Leg} = \frac{7b^2}{24}$$

The length of any middle leg is equal to the distance between a (approximately) random point (the destination of the most recently delivered customer) and the nearest destination of anyone of the customers who still remain on the vehicle. Computing the expected value and variance of this distance is a far more complicated and tedious problem due to the effects of the region's boundaries. We pursued two different approaches for approximating these quantities using: (a) a Crofton Approximation (Technical Section 3.1.T3) that computes the expected distance and variance of the distance between a random point and the closest of N ($N = 1, 2, 3, \ldots, c-1$) other random points on a linear segment using Crofton's Method[7] and then treats the distances in the horizontal and vertical directions, as if they are independent; and (b) a Center Approximation (Technical Section 3.1.T4) that relies on computing the expected value and variance of the distance between the center of the rectangular service region and the closest of N ($N = 1, 2, 3, \ldots, c-1$) random points in the rectangle.

We then tested the analytical expressions derived through (a) and (b) by means of an extensive series of numerical experiments, described in Technical Section 3.1.T5. The experiments indicated that the expressions performed equally well, but we have chosen to use the Crofton Approximation henceforth because of its simpler form. We have also used linear regression models to correct the Crofton and Center expressions, so they fit better with the numerical observations. It was found that, again, both of the corrected expressions perform roughly equally and will use henceforth the Crofton Approximation with/without the regression correction because of its simpler form.

In conclusion, our best estimates for the first and second moments of the length of a middle leg of the delivery tour, given that N customers remain to be delivered, are given by the following expressions:

$$s_{middle,N} \approx s_{N,Crofton\,Approx} = \frac{(N+3)(a+b)}{2(N+1)(N+2)} \tag{13}$$

$$SQ_{middle,N} \approx SQ_{N,Crofton\,Approx} = \frac{(N+7)(a^2+b^2)}{2(N+1)(N+2)(N+3)} + 2(\frac{N+3}{2(N+1)(N+2)})^2 ab \tag{14}$$

After correcting these expressions through regression, they become:

$$s_{middle,N} \approx s_{N,Crofton\,Approx} \approx (1.13047 + 0.099945N) \cdot \frac{(N+3)(a+b)}{2(N+1)(N+2)} \tag{15}$$

$$SQ_{middle,N} \approx SQ_{N,Crofton\,Approx}$$

$$\approx (0.525751 + 0.372122N) \cdot (\frac{(N+7)(a^2+b^2)}{2(N+1)(N+2)(N+3)}$$

$$+ 2\left(\frac{N+3}{2(N+1)(N+2)}\right)^2 ab \tag{16}$$

The detailed mathematical derivation of (13) and (14) is in Technical Section 3.1.T3 and of (15) and (16) in Technical Section 3.1.T5.

3.1.T3*: Expectation and variance approximation for middle leg using Crofton's method.

We need to find the distance between a (approximately) random point (the destination of the most recently delivered customer) and the nearest destination of anyone of the customers who still remain on the vehicle, using Crofton's method:

At first, we define three problems:

Problem 0: Node A is uniformly distributed from 0 to a, N other nodes are identically, independently and uniformly distributed from 0 to a. Let $D_{0,N}$ denotes the distance between Node A and the closest node of N other nodes, define:

$$s_{0,N} = Expectation(D_{0,N}), VAR_{0,N} = Variance(D_{0,N}), SQ_{0,N} = Expectation(D_{0,N}^2)$$

Problem 1: Node A is uniformly distributed from 0 to a, Node B is located on point 0, N other nodes are identically, independently and uniformly distributed from 0 to a. Let $D_{1,N}$ denotes the distance between Node A and the closest node of Node B and N other nodes, define:

$$s_{1,N} = Expectation(D_{1,N}), VAR_{1,N} = Variance(D_{1,N}), SQ_{1,N} = Expectation(D_{1,N}^2)$$

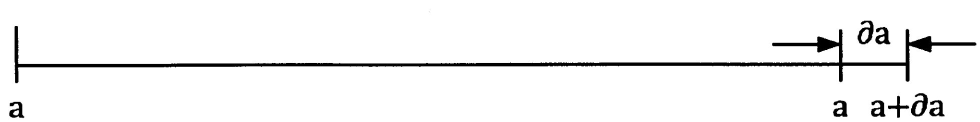
Problem 2: Node A is uniformly distributed from 0 to a, Node B is located on point 0, Node C is located on point a, N other nodes are identically, independently and uniformly distributed from 0 to a. Let $D_{2,N}$ denotes the distance between Node A and the closest node of Node B, Node C and N other nodes, define:

$$s_{2,N} = Expectation(D_{2,N}), VAR_{2,N} = Variance(D_{2,N}), SQ_{2,N} = Expectation(D_{2,N}^2)$$

What we need is the results of *Problem 0*, whose deducing process needs the results of *Problem 1* as the boundary condition, while additionally the deducing process of *Problem 1* needs the results of *Problem 2* as the boundary condition.

We study and analyze the $s_{0,N}$, $s_{1,N}$ and $s_{2,N}$ as follows:

For *Problem 0*, we add to the interval $[0, a]$ an increment of length ∂a, as illustrated in the following figure. We now consider the problem in which Node A and N other nodes are independent and distributed in the same way as before, but over the larger interval $[0, a + \partial a]$. Then $s_{0,N}$ becomes $s_{0,N} + \partial s_{0,N}$. Consider the following four mutually exclusive events:



E_1: *Node A and all other N nodes lie in* $[0, a]$, $P(E_1) = (\frac{a}{a + \partial a})^{N+1}$

E_2: *Node A lies in* $[a, a + \partial a]$, *all other N nodes lie in* $[0, a]$, $P(E_2) = \frac{\partial a}{a + \partial a} \cdot (\frac{a}{a + \partial a})^{N}$

E_3: *Node A lies in* $[0, a]$, $N - 1$ *of N nodes lie in* $[0, a]$, *one of N nodes lies in* $[a, a + \partial a]$,

$$P(E_3) = \frac{a}{a + \partial a} \cdot \frac{N \cdot \partial a}{a + \partial a} \cdot (\frac{a}{a + \partial a})^{N-1}$$

E_4: *all other events,* $P(E_4) = O(\partial a^2) = o(\partial a)$

Under condition of E_1, all nodes are distributed in [0, a], like the situation before adding ∂a:

$$E[D_{0,N}|E_1] = s_{0,N}$$

Under condition of E_2, Node A is on the end, while other *N* nodes are distributed in [0, a], so the distance we need is like the first-order statistics:

$$E[D_{0,N}|E_2] = \frac{a}{N + 1}$$

Under condition of E_3, one node lies on the end, while Node A and other $N - 1$ nodes are distributed in $[0, a]$, like the situation of *Problem1* with parameter $N - 1$:

$$E[D_{0,N}|E_3] = s_{1,N-1}$$

Under condition of E_4, since $P(E_4) = O(\partial a^2) = o(\partial a)$, we do not care $E[D_{0,N}|E_4]$.

Now $s_{0,N} + \partial s_{0,N}$ can be written as the weighted sum of four conditional expected values, the weights being the appropriate probabilities:

$$s_{0,N} + \partial s_{0,N} = \sum_{i=1}^{4} E[D_{0,N}|E_i]P(E_i)$$

Substituting, we have

$$s_{0,N} + \partial s_{0,N} = (\frac{a}{a+\partial a})^{N+1} \cdot s_{0,N} + \frac{\partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N} \cdot \frac{a}{N+1} + \frac{a}{a+\partial a} \cdot \frac{N \cdot \partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N-1} \cdot s_{1,N-1} \quad (3.1.T3.1)$$

This will yield a differential equation group with respect to $s_{0,N}$.

Similarly, for *Problem 1*, we add to the interval [0, a] an increment of length ∂a, as illustrated in the following figure. We now consider the problem in which Node A and N other nodes are independent and distributed in the same way as before, but over the larger interval $[0,\ a+\partial a]$. Then $s_{1,N}$ becomes $s_{1,N} + \partial s_{1,N}$. Node B is on the left end 0. Consider the following four mutually exclusive events:

E_1: *Node A and all other N nodes lie in* $[0, a]$, $\quad P(E_1) = (\frac{a}{a+\partial a})^{N+1}$

E_2: *Node A lies in* $[a, a+\partial a]$, *all other N nodes lie in* $[0, a]$, $\quad P(E_2) = \frac{\partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N}$

E_3: *Node A lies in* $[0, a]$, $N-1$ *of N nodes lie in* $[0, a]$, *one of N nodes lies in* $[a, a+\partial a]$,

$$P(E_3) = \frac{a}{a+\partial a} \cdot \frac{N \cdot \partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N-1}$$

E_4: *all other events,* $\quad P(E_4) = O(\partial a^2) = o(\partial a)$

Under condition of E_1, all nodes are distributed in $[0, a]$, like the situation before adding ∂a:

$$E[D_{1,N} | E_1] = s_{1,N}$$

Under condition of E_2, Node A is on the right end, while other N nodes are distributed in [0, a], so the distance we need is like the first-order statistics:

$$E[D_{1,N} | E_2] = \frac{a}{N+1}$$

Under condition of E_3, Node B lies on the left end, one other node lies on the right end, and Node A and other $N-1$ nodes are distributed in $[0,a]$, like the situation of *Problem2* with parameter $N-1$:

$$E[D_{1,N}|E_3] = s_{2,N-1}$$

Under condition of E_4, since $P(E_4) = O(\partial a^2) = o(\partial a)$, we do not care $E[D_{1,N}|E_4]$.

Now $s_{1,N} + \partial s_{1,N}$ can be written as the weighted sum of four conditional expected values, the weights being the appropriate probabilities:

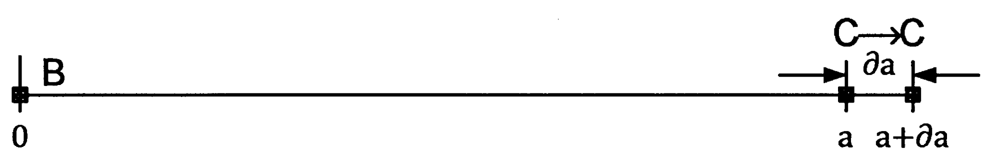
$$s_{1,N} + \partial s_{1,N} = \sum_{i=1}^{4} E[D_{1,N}|E_i]P(E_i)$$

Substituting, we have

$$s_{1,N} + \partial s_{1,N} = (\frac{a}{a+\partial a})^{N+1} \cdot s_{1,N} + \frac{\partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N} \cdot \frac{a}{N+1} + \frac{a}{a+\partial a} \cdot \frac{N \cdot \partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N-1} \cdot s_{2,N-1} \tag{3.1.T3.2}$$

This will yield a differential equation group with respect to $s_{1,N}$.

Similarly, For *Problem 2*, we add to the interval $[0, a]$ an increment of length ∂a, as illustrated in the following figure. We now consider the problem in which Node A and N other nodes are independent and distributed in the same way as before, but over the larger interval $[0, a + \partial a]$. Then $s_{2,N}$ becomes $s_{2,N} + \partial s_{2,N}$. Node B is on the left end 0, Node C move from a to $a + \partial a$. Consider the following four mutually exclusive events:



E_1: *Node A and all other N nodes lie in* $[0,a]$, $\quad P(E_1) = (\frac{a}{a+\partial a})^{N+1}$

E_2: *Node A lies in* $[a, a+\partial a]$, *all other N nodes lie in* $[0,a]$, $\quad P(E_2) = \frac{\partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N}$

E_3: *Node A lies in* $[0, a]$, $N-1$ *of N nodes lie in* $[0, a]$, *one of N nodes lies in* $[a, a+\partial a]$,

$$P(E_3) = \frac{a}{a+\partial a} \cdot \frac{N \cdot \partial a}{a+\partial a} \cdot \left(\frac{a}{a+\partial a}\right)^{N-1}$$

E_4: *all other events*, $$P(E_4) = O(\partial a^2) = o(\partial a)$$

Under condition of E_1, all nodes are distributed in $[0, a]$, $s_{2,N}$ will increase ∂a if and only if (1) Node A is the most right node among the $N+1$ uniformly distributed nodes, (2) the distance between Node A and Node C is smaller than the distance between Node A and its nearest node on the left, and the probability is $1/2 \times 1/(N+1) = 1/(2(N+1))$. Therefore,

$$E[D_{2,N}|E_1] = s_{2,N} + \frac{1}{2(N+1)} \cdot \partial a$$

Under condition of E_2, Node A is on the right end, while other N nodes are distributed in $[0, a]$, so the distance between Node A and Node C is $O(\partial a)$:

$$E[D_{2,N}|E_2] = O(\partial a)$$

Under condition of E_3, Node B lies on the left end, Node C as well as one other node lie on the right end, and Node A and other $N-1$ nodes are distributed in $[0, a]$, like the situation of *Problem2* with parameter $N-1$:

$$E[D_{2,N}|E_3] = s_{2,N-1}$$

Under condition of E_4, since $P(E_4) = O(\partial a^2) = o(\partial a)$, we do not care $E[D_{2,N}|E_4]$.

Now $s_{2,N} + \partial s_{2,N}$ can be written as the weighted sum of four conditional expected values, the weights being the appropriate probabilities:

$$s_{2,N} + \partial s_{2,N} = \sum_{i=1}^{4} E[D_{2,N}|E_i] P(E_i)$$

Substituting, we have

$$s_{2,N} + \partial s_{2,N} = (\frac{a}{a+\partial a})^{N+1} \cdot (s_{2,N+} \frac{1}{2(N+1)} \cdot \partial a) + \frac{\partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N} \cdot O(\partial a) + \frac{a}{a+\partial a} \cdot \frac{N \cdot \partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N-1} \cdot s_{2,N-1} \quad (3.1.T3.3)$$

This will yield a differential equation group with respect to $s_{2,N}$.

We know the boundary condition of $s_{2,N}$, that is $s_{2,0} = a/4$, then solve the third differential equation group, we obtain:

$$s_{2,N} = \frac{a}{2(N+2)}$$

Taking it as the boundary condition of the second differential equation group, we obtain:

$$s_{1,N} = \frac{a}{2(N+1)}$$

Taking it as the boundary condition of the first differential equation group, we obtain:

$$s_{0,N} = \frac{(N+3)a}{2(N+1)(N+2)} \quad (3.1.T3.4)$$

In order to obtain $SQ_{0,N}$, $SQ_{1,N}$ and $SQ_{2,N}$, we construct the similar differential equation groups as before, using Crofton's Method:

Problem 0:

$$SQ_{0,N} + \partial SQ_{0,N} = (\frac{a}{a+\partial a})^{N+1} \cdot SQ_{0,N} + \frac{\partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N} \cdot \frac{2a^2}{(N+1)(N+2)} + \frac{a}{a+\partial a} \cdot \frac{N \cdot \partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N-1} \cdot SQ_{1,N-1} \quad (3.1.T3.5)$$

Problem 1:

$$SQ_{1,N} + \partial SQ_{1,N} = (\frac{a}{a+\partial a})^{N+1} \cdot SQ_{1,N} + \frac{\partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N} \cdot \frac{2a^2}{(N+1)(N+2)} + \frac{a}{a+\partial a} \cdot \frac{N \cdot \partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N-1} \cdot SQ_{2,N-1} \quad (3.1.T3.6)$$

Problem 2:

$$SQ_{2,N} + \partial SQ_{2,N} = (\frac{a}{a+\partial a})^{N+1} \cdot (SQ_{2,N+}\frac{1}{2(N+1)} \cdot 2 \cdot \frac{a}{2(N+2)} \cdot \partial a) + \frac{\partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N}$$

$$\cdot O(\partial a) + \frac{a}{a+\partial a} \cdot \frac{N \cdot \partial a}{a+\partial a} \cdot (\frac{a}{a+\partial a})^{N-1} \cdot SQ_{2,N-1} \tag{3.1.T3.7}$$

We know the boundary condition of $SQ_{2,N}$, that is $SQ_{2,0} = a^2/12$, then solve the third differential equation group, we obtain:

$$SQ_{2,N} = \frac{a^2}{2(N+2)(N+3)}$$

Taking it as the boundary condition of the second differential equation group, we obtain:

$$SQ_{1,N} = \frac{(N+4)a^2}{2(N+1)(N+2)(N+3)}$$

Taking it as the boundary condition of the first differential equation group, we obtain:

$$SQ_{0,N} = \frac{(N+7)a^2}{2(N+1)(N+2)(N+3)} \tag{3.1.T3.8}$$

From above, we obtain the exact analytical solution to the one-dimension problem:

$$s_{0,N} = \frac{(N+3)a}{2(N+1)(N+2)}$$

$$SQ_{0,N} = \frac{(N+7)a^2}{2(N+1)(N+2)(N+3)}$$

$$VAR_{0,N} = Variance = SQ_{0,N} - s_{0,N}^2 = \frac{(N+7)a^2}{2(N+1)(N+2)(N+3)} - \left(\frac{(N+3)a}{2(N+1)(N+2)}\right)^2$$

$$= \frac{N^3 + 11N^2 + 19N + 1}{4(N+1)^2(N+2)^2(N+3)}a^2$$

Therefore, in the original two-dimension problem, if we assume the distance traveled in *X* direction and the distance traveled in *Y* direction are independent, then:

$$s_{N,Crofton\ Approx} = \frac{(N+3)(a+b)}{2(N+1)(N+2)} \tag{3.1.T3.9}$$

$$SQ_{N,Crofton\ Approx} = \frac{(N+7)(a^2+b^2)}{2(N+1)(N+2)(N+3)} + 2(\frac{N+3}{2(N+1)(N+2)})^2 ab \tag{3.1.T3.10}$$

3.1.T4*: Expectation and variance approximation for middle leg using Center approximation method

We need to find the distance between the center of the rectangle and the closest of N random points in the rectangle. It is similar to the first leg in the route, from the analysis in Technical Section 3.1.T1, we obtain:

When the region is rectangle, and $a > b$,

$$s_{N,Center\ Approx} =$$

$$= \frac{1}{4}\left(\frac{2^{-N}a^{-N}((2a-b)^{1+N} - b^{1+N})}{1+c} + \frac{2^{1-N}b(\frac{b}{a})^N}{1+2c} + 2b\text{Hypergeometric2F1}[\frac{1}{2}, -N, \frac{3}{2}, \frac{b}{2a}]\right) \quad (3.1.T4.1)$$

$$SQ_{N,Center\ Approx}$$

$$= 2^{-3-N}\left(\frac{2b(2^{1+N}a - (2-\frac{b}{a})^N(2a+bN))}{1+N} + \frac{2(\frac{b}{a})^N(b^2+2ab(1+N)+a^2(1+3N+2N^2))}{1+3N+2N^2} + \frac{1}{(1+N)(2+N)}a^{-N}((2a-b)^N((2a+b)^2+4b(a+b)N+2b^2N^2) - b^c(b^2+2ab(2+N)+2a^2(1+N)(2+N)))\right) \quad (3.1.T4.2)$$

When the region is square, i.e., $a = b$,

$$s_{N,Center\ Approx} = \frac{2^{-2-N}b(1+(1+2N)\text{Hypergeometric2F1}[1, -1-N, \frac{1}{2}, -1])}{1+3N+2N^2} \quad (3.1.T4.3)$$

$$SQ_{N,Center\ Approx} = \frac{2^{-1-N}b^2(1+2^N+2^{1+N}N)}{1+3N+2N^2} \quad (3.1.T4.4)$$

where Hypergeometric2F1 $= F_1(a,b;c;z) = \sum_{k=0}^{\infty}(a)_k(b)_k/(c)_k z^k/k!$, $(a)_k = a(a+1)\dots(a+k-1) = \Gamma(a+k)/\Gamma(a)$, $(b)_k = b(b+1)\dots(b+k-1) = \Gamma(b+k)/\Gamma(b)$, $(c)_k = c(c+1)\dots(c+k-1) = \Gamma(c+k)/\Gamma(c)$.

3.1.T5*: Regression correction for middle leg approximations

We develop a numerical experiment to simulate the real distance expectation and second moment of the middle leg, then compare the real (simulated) value to the approximated values obtained with the two methods described before. In the experiment, we assume a square service region with $a = b = 150$.

The real (simulated) values of distance expectation are shown in the table bellow, along with the corresponding Crofton Approximation and Center Approximation:

	Expectation: S (sec)						
N	**Simulation**		**Crofton**	Simulation/Crofton		**Center**	Simulation/Center
1	99.97		100.00	0.9997		100.00	0.9997
2	73.26		62.50	1.1722		57.50	1.2741
3	60.26		45.00	1.3391		48.75	1.2361
4	52.19		35.00	1.4912		43.18	1.2086
5	46.66		28.57	1.6330		39.21	1.1899
6	42.50		24.11	1.7629		36.18	1.1747
7	39.20		20.83	1.8816		33.76	1.1611
8	36.55		18.33	1.9938		31.77	1.1505
9	34.39		16.36	2.1014		30.10	1.1424
10	32.56		14.77	2.2042		28.67	1.1359
11	30.94		13.46	2.2985		27.42	1.1285
12	29.58		12.36	2.3923		26.32	1.1236
13	28.36		11.43	2.4812		25.35	1.1187
14	27.26		10.63	2.5657		24.47	1.1139
15	26.29		9.93	2.6489		23.68	1.1102
16	25.42		9.31	2.7296		22.97	1.1069
17	24.64		8.77	2.8095		22.31	1.1046
18	23.87		8.29	2.8796		21.71	1.0997
19	23.23		7.86	2.9561		21.15	1.0982
20	22.60		7.47	3.0266		20.63	1.0953

Table 1: Simulated and approximate middle leg expectation

We run linear regression with *points number* $N(>=2)$ as independent variable, *Simulation/Crofton* and *Simulation/Center* as induced variables, to obtain the following linear approximations:

$Simulation/Crofton = 1.13047 + 0.099945N$, with $R^2 = 0.992613$;

$Simulation/Center = 1.236623 - 0.00824N$, with $R^2 = -0.92095$.

Therefore, we use the following approximations with linear regression correction:

$$s \approx s_{N,Crofton\ Approx} \cdot \frac{Simulation}{Crofton} = (c_{crofton} + d_{crofton}N) \cdot \frac{(N+3)(a+b)}{2(N+1)(N+2)} \quad (3.1.T5.1)$$

Where $c_{crofton} = 1.13047, d_{crofton} = 0.099945$;

$$s \approx s_{N,Center\ Approx} \cdot \frac{Simulation}{Center} = (c_{center} + d_{center}N) \cdot \frac{2^{-2-N}b(1+(1+2N)\text{Hypergeometric2F1}[1,-1-N,\frac{1}{2},-1])}{1+3N+2N^2} \quad (3.1.T5.2)$$

Where $c_{center} = 1.236623, d_{center} = -0.00824$.

The approximations after regression correction are as follows:

	Expectation: S (sec)						
N	***Simulation***		***Crofton Approx.***	*Error*		***Center Approx.***	*Error*
2	73.26		83.15	13.49%		70.16	-4.24%
3	60.26		64.36	6.81%		59.08	-1.95%
4	52.19		53.56	2.62%		51.98	-0.41%
5	46.66		46.58	-0.17%		46.87	0.46%
6	42.50		41.71	-1.86%		42.95	1.06%
7	39.20		38.13	-2.74%		39.80	1.53%
8	36.55		35.38	-3.20%		37.20	1.76%
9	34.39		33.22	-3.40%		34.99	1.75%
10	32.56		31.46	-3.37%		33.09	1.61%
11	30.94		30.02	-2.99%		31.42	1.55%
12	29.58		28.80	-2.61%		29.95	1.26%
13	28.36		27.77	-2.07%		28.63	0.96%
14	27.26		26.88	-1.40%		27.44	0.66%
15	26.29		26.10	-0.73%		26.36	0.25%
16	25.42		25.42	0.00%		25.37	-0.20%
17	24.64		24.82	0.71%		24.46	-0.74%
18	23.87		24.28	1.73%		23.62	-1.04%
19	23.23		23.80	2.48%		22.84	-1.66%
20	22.60		23.37	3.40%		22.12	-2.15%

Table 2: Middle leg expectation approximation after regression correction

Similarly, the real (simulated) values of distance second moment are shown in the table bellow, along with the corresponding Crofton Approximation and Center Approximation:

	Square Expectation: SQ (sec^2)				
N	**Simulation**	**Crofton**	*Simulation/Crofton*	**Center**	*Simulation/Center*
1	12504.53	12500.00	1.0004	6562.50	1.9055
2	6901.71	5328.13	1.2953	3937.50	1.7528
3	4736.67	2887.50	1.6404	2862.72	1.6546
4	3590.87	1791.07	2.0049	2265.63	1.5849
5	2880.49	1211.73	2.3772	1880.33	1.5319
6	2391.94	870.93	2.7464	1609.07	1.4865
7	2044.55	654.51	3.1238	1406.98	1.4531
8	1779.47	508.96	3.4963	1250.29	1.4232
9	1572.48	406.61	3.8673	1125.12	1.3976
10	1409.21	332.02	4.2444	1022.77	1.3778
11	1277.72	276.05	4.6287	937.52	1.3629
12	1164.55	233.01	4.9979	865.39	1.3457
13	1070.78	199.23	5.3745	803.58	1.3325
14	988.58	172.25	5.7391	750.00	1.3181
15	919.25	150.37	6.1132	703.13	1.3074
16	859.61	132.38	6.4934	661.77	1.2990
17	803.69	117.42	6.8445	625.00	1.2859
18	758.26	104.85	7.2321	592.11	1.2806
19	714.45	94.18	7.5861	562.50	1.2701
20	676.82	85.05	7.9576	535.71	1.2634

Table 3: Simulated and approximate middle leg second moment

We run linear regression with *points number* $N(>=2)$ as independent variable, *Simulation/Crofton* and *Simulation/Center* as induced variables, and obtain the following linear approximations:

$Simulation/Crofton = 0.525751 + 0.372122N$, with $R^2 = 0.999988$;

$Simulation/Center = 1.659275 - 0.02296N$, with $R^2 = -0.92835$.

Therefore, we use the following approximations with linear regression correction:

$$SQ \approx SQ_{N,Crofton\ Approx} \cdot \frac{Simulation}{Crofton}$$

$$= \left(c_{crofton,q} + d_{crofton,q}N\right) \cdot \left(\frac{(N+7)(a^2+b^2)}{2(N+1)(N+2)(N+3)} + 2\left(\frac{N+3}{2(N+1)(N+2)}\right)^2 ab\right) \quad (3.1.T5.3)$$

Where $c_{crofton} = 0.525751, d_{crofton} = 0.372122;$

$$SQ \approx SQ_{N,Center\ Approx} \cdot \frac{Simulation}{Center} = (c_{center} + d_{center}N) \cdot \frac{2^{-1-N}b^2(1+2^N+2^{1+N}N)}{1+3N+2N^2}$$

Where $c_{center} = 1.659275, d_{center} = -0.02296.$

The approximations after regression correction are as follows:

	Square Expectation: SQ (sec^2)				
N	**Simulation**	**Crofton Approx.**	Error	**Center Approx.**	Error
2	6901.71	6766.69	-1.96%	6352.61	-7.96%
3	4736.67	4741.61	0.10%	4552.89	-3.88%
4	3590.87	3607.65	0.47%	3551.25	-1.10%
5	2880.49	2891.64	0.39%	2904.15	0.82%
6	2391.94	2402.46	0.44%	2448.26	2.35%
7	2044.55	2049.02	0.22%	2108.47	3.13%
8	1779.47	1782.76	0.19%	1844.95	3.68%
9	1572.48	1575.56	0.20%	1634.41	3.94%
10	1409.21	1410.07	0.06%	1462.27	3.77%
11	1277.72	1275.08	-0.21%	1318.85	3.22%
12	1164.55	1163.01	-0.13%	1197.52	2.83%
13	1070.78	1068.56	-0.21%	1093.53	2.12%
14	988.58	987.96	-0.06%	1003.41	1.50%
15	919.25	918.40	-0.09%	924.55	0.58%
16	859.61	857.80	-0.21%	854.98	-0.54%
17	803.69	804.55	0.11%	793.13	-1.31%
18	758.26	757.40	-0.11%	737.79	-2.70%
19	714.45	715.39	0.13%	687.99	-3.70%
20	676.82	677.72	0.13%	642.93	-5.01%

Table 4: Middle leg second moment approximation after regression correction

3.2 Completion of the Queuing Model

In this subsection, we incorporate the results of the above Chapter 3.1 into the previously (Chapter 2) derived results for the Unit-Capacity queuing model to obtain approximations of system performance for the General ($c > 1$) Capacity case. Specifically, we use the expressions for the length and duration of customer delivery tours when $c > 1$, to estimate the service times for the General Capacity model and use these estimates in the various expressions for the expected waiting time until boarding a vehicle that were obtained in Chapter 2.2.2 under the cyclic assignment policy. As was demonstrated in Chapter 2.3, these latter expressions approximate best the observed (through simulation) system performance.

For the case of a General distribution for the size of customer batches and of General service times the strict cyclic upper bound [cf. expression (7)] and the approximate cyclic upper bound [cf. expression (9)] for the waiting time until boarding a vehicle (see Technical Section 3.2.T) for details) is then given by:

$$W_{Board,strict} \leq \frac{4mE^2(\xi_E)E(S_E{}^2) - 4E^2(S_E)E^3(\xi_E) + 4mhE(S_E)E(\xi_E{}^2) + m^3hE(S_E) - 4m^2hE(S_E)E(\xi_E)}{8m\big(mh - E(\xi_E)E(S_E)\big)E(\xi_E)} \tag{17}$$

$$W_{Board,approx} \approx \frac{\rho(C_a^2 + C_{T_E}^2)E(T_E)}{2(1-\rho)} \cdot \exp\left[-\frac{2(1-\rho)(1-C_a^2)^2}{3\rho(C_a^2 + C_{T_E}^2)}\right] + \frac{E(S_E)\cdot\big(4E(\xi_E{}^2) + m^2 - 4mE(\xi_E)\big)}{8mE(\xi_E)} \tag{18}$$

When the size of customer batches has a Poisson distribution, and the duration of the delivery service tour is approximated through Crofton's method (without using the regression correction), the various terms of (17) and (18) above take the following values:

$$E(\xi_E) \approx \frac{E(\xi)}{c}, \qquad VAR(\xi_E) \approx \frac{4VAR(\xi) + c^2}{4c^2}, \qquad C_a^2 = 0, \qquad \rho = \frac{E(S_E)E(\xi_E)}{mh},$$

$$E(T_E) = \frac{E(S_E)E(\xi_E)}{m}, \qquad C_{T_E}^2 = \frac{4mE(\xi_E)Var(S_E) + 4E^2(S_E)Var(\xi_E) + E^2(S_E)m^2}{4E^2(\xi_E)E^2(S_E)},$$

Hypergeometric2F1 $= F_1(a,b;c;z) = \sum_{k=0}^{\infty}(a)_k\,(b)_k/(c)_k\, z^k/k!$,

$(a)_k = a(a+1)\dots(a+k-1) = \Gamma(a+k)/\Gamma(a)$, $(b)_k = b(b+1)\dots(b+k-1) = \Gamma(b+k)/\Gamma(b)$, $(c)_k = c(c+1)\dots(c+k-1) = \Gamma(c+k)/\Gamma(c)$

$$E(S_E) = \frac{2^{-2-c}b(1+(1+2c)\text{Hypergeometric2F1}[1,-1-c,\frac{1}{2},-1])}{1+3c+2c^2} + \sum_{i=1}^{c-1}\frac{(i+3)b}{(i+1)(i+2)} + \frac{b}{2}$$

$$E(S_E{}^2) = \frac{2^{-1-c}b^2(1+2^c+2^{1+c}c)}{1+3c+2c^2} - \frac{2^{-4-2c}b^2(1+(1+2c)\text{Hypergeometric2F1}[1,-1-c,\frac{1}{2},-1])^2}{(1+3c+2c^2)^2} + \sum_{i=1}^{c-1}\frac{i^3+11i^2+19i+1}{2(i+1)^2(i+2)^2(i+3)}b^2 + \frac{7b^2}{24}$$

Note that in (17) and (18) we have used the notation $W_{Board,strict}$ and $W_{Board,approx}$ for the expected waiting time until a customer will board a vehicle, while in (7) and (9) we used the notation W in (7) and (9) for the same quantity. This is because we also want to introduce here another quantity, W_{Riding}, which is defined as the expected time a customer will spend riding on the vehicle before being delivered to his destination. Considering the riding component of the trip, the total expected time from the instant a customer arrives at the rail station until she is delivered at her destination is given by

$$W_{Delivered} = W_{Board} + W_{Riding}$$

where

$$W_{Riding} = \frac{2^{-2-c}b(1+(1+2c)\text{Hypergeometric2F1}[1,-1-c,\frac{1}{2},-1])}{1+3c+2c^2} + \frac{c-1}{c+1}b$$

as shown in Technical Section 3.2.T.

3.2.T*: Upper bound and approximation of the General-Capacity, Multi-Vehicle LMP under cyclic assignment policy

Average waiting time until boarding and average waiting time until delivery for General customer batch size and General service time distributions:

$$W_{Board,strict} \le \frac{4mE^2(\xi_E)E(S_E{}^2) - 4E^2(S_E)E^3(\xi_E) + 4mhE(S_E)E(\xi_E{}^2) + m^3hE(S_E) - 4m^2hE(S_E)E(\xi_E)}{8m(mh - E(\xi_E)E(S_E))E(\xi_E)}$$

$$W_{Board,strict} \le \frac{\rho(C_a^2 + C_{T_E}^2)E(T_E)}{2(1-\rho)} \cdot \exp\left[-\frac{2(1-\rho)(1-C_a^2)^2}{3\rho(C_a^2 + C_{T_E}^2)}\right] + \frac{E(S_E)\cdot\left(4E(\xi_E{}^2) + m^2 - 4mE(\xi_E)\right)}{8mE(\xi_E)}$$

(3.2.T.2)

The average waiting time until delivery to the final destination is equal to the sum of average waiting time until boarding the vehicle and the average riding time on road: $W_{Delivered} = W_{Board} + W_{Riding}$.

For the case of Poisson passenger size and Crofton's method service time approximation without regression correction:

The service time expectation: (Using Crofton Approximation, without regression, $a = b$)

$$E(S_E) = Expectation(Total\ Service\ Time) = s_{First\ Leg} + s_{middle,c-1} + s_{middle,c-2} + \cdots + s_{middle,1} + s_{Last\ Leg}$$

$$= \frac{2^{-2-c}b(1 + (1+2c)\text{Hypergeometric2F1}[1, -1-c, \frac{1}{2}, -1])}{1 + 3c + 2c^2} + \sum_{i=1}^{c-1}\frac{(i+3)b}{(i+1)(i+2)} + \frac{b}{2}$$

The service time variance:

$$
\begin{aligned}
VAR_E &= Variance(Total\ Service\ Time) \\
&\approx VAR_{First\ Leg} + VAR_{middle,c-1} + VAR_{middle,c-2} + \cdots + VAR_{middle,1} \\
&+ VAR_{Last\ Leg} \\
&= \frac{2^{-1-c}b^2(1+2^c+2^{1+c}c)}{1+3c+2c^2} \\
&- \frac{2^{-4-2c}b^2(1+(1+2c)\text{Hypergeometric2F1}[1,-1-c,\frac{1}{2},-1])^2}{(1+3c+2c^2)^2} \\
&+ \sum_{i=1}^{c-1} \frac{i^3+11i^2+19i+1}{2(i+1)^2(i+2)^2(i+3)} b^2 + \frac{7b^2}{24}
\end{aligned} \tag{3.2.T.4}
$$

The expectation of service time second moment:

$$
\begin{aligned}
E({S_E}^2) &= SQ_E = VAR_E + (E(S_E))^2 \\
&\approx \frac{2^{-1-c}b^2(1+2^c+2^{1+c}c)}{1+3c+2c^2} \\
&- \frac{2^{-4-2c}b^2(1+(1+2c)\text{Hypergeometric2F1}[1,-1-c,\frac{1}{2},-1])^2}{(1+3c+2c^2)^2} \\
&+ \sum_{i=1}^{c-1} \frac{i^3+11i^2+19i+1}{2(i+1)^2(i+2)^2(i+3)} b^2 + \frac{7b^2}{24} \\
&+ (\frac{2^{-2-c}b(1+(1+2c)\text{Hypergeometric2F1}[1,-1-c,\frac{1}{2},-1])}{1+3c+2c^2} \\
&+ \sum_{i=1}^{c-1} \frac{(i+3)b}{(i+1)(i+2)} + \frac{b}{2})^2
\end{aligned} \tag{3.2.T.5}
$$

$$
C_{S_E}^2 = \frac{VAR(S_E)}{E^2(S_E)} = \frac{E^2(S_E) - E({S_E}^2)}{E^2(S_E)} = \frac{E({S_E}^2)}{E^2(S_E)} - 1
$$

$$
C_a^2 = 0
$$

$$
\rho = \frac{E(S_E)E(\xi_E)}{mh}
$$

$$
E(\xi_E) \approx \frac{E(\xi)}{n}
$$

$$
E({\xi_E}^2) = E^2(\xi_E) + VAR(\xi_E) \approx \left(\frac{E(\xi)}{c}\right)^2 + \frac{4VAR(\xi) + c^2}{4c^2}
$$

$$E(T_E)=\frac{E(S_E)E(\xi_E)}{m}$$

$$C_{t_E}^2=\frac{E(N_E)Var(S_E)+E^2(S_E)Var(N_E)}{E^2(N_E)E^2(S_E)}=\frac{E\left(\frac{\xi_E}{m}\right)Var(S_E)+E^2(S_E)Var\left(\frac{\xi_E}{m}\right)}{E^2\left(\frac{\xi_E}{m}\right)E^2(S_E)}$$

$$=\frac{E(\xi_E)Var(S_E)/m+E^2(S_E)\frac{4Var(\xi_E)+m^2}{4m^2}}{E^2(\xi_E)E^2(S_E)/m^2}$$

$$=\frac{4mE(\xi_E)Var(S_E)+4E^2(S_E)Var(\xi_E)+E^2(S_E)m^2}{4E^2(\xi_E)E^2(S_E)}$$

Average riding time on road without regression correction:

For the first customer in one loop: $Expectation(Time\ on\ Road)_1 = s_{First\ Leg}$

For the second customer in one loop: $Expectation(Time\ on\ Road)_2 = s_{First\ Leg} + s_{middle,c-1}$

……

For the last customer in one loop: $Expectation(Time\ on\ Road)_c = s_{First\ Leg} + s_{middle,c-1} + s_{middle,c-2} + \cdots + s_{middle,1}$

Therefore, the average riding time on road is

$$W_{Riding}=\frac{\sum_{i=1}^{c} Expectation(Time\ on\ Road)_i}{c}=s_{First\ Leg}+\sum_{i=1}^{c-1}\frac{i}{c}s_{middle,i}$$

$$=\frac{2^{-2-c}b(1+(1+2c)\text{Hypergeometric2F1}[1,-1-c,\frac{1}{2},-1])}{1+3c+2c^2}$$

$$+\sum_{i=1}^{c-1}\frac{i(i+3)b}{c(i+1)(i+2)}$$

$$=\frac{2^{-2-c}b(1+(1+2c)\text{Hypergeometric2F1}[1,-1-c,\frac{1}{2},-1])}{1+3c+2c^2}+\frac{c-1}{c+1}b$$

Similarly, for the case of Poisson passenger size and Crofton' method service time approximation with regression correction:

The service time expectation: (Using Crofton Approximation, with regression, $a = b$)

$$E(S_E) = Expectation(Total\ Service\ Time)$$
$$= s_{First\ Leg} + s_{middle,c-1} + s_{middle,c-2} + \cdots + s_{middle,1} + s_{Last\ Leg}$$
$$= \frac{2^{-2-c} b\left(1 + (1+2c)\text{Hypergeometric2F1}\left[1, -1-c, \frac{1}{2}, -1\right]\right)}{1 + 3c + 2c^2}$$
$$+ \sum_{i=1}^{c-1} \frac{(c_1 + c_2 i)(i+3)b}{(i+1)(i+2)} + \frac{b}{2} \tag{3.2.T.7}$$

Where $c_1 = 1.13; c_2 = 0.999$.

The service time variance:

$$SQ_n = (c_3 + c_4 N) \cdot \left(\frac{(N+7)(a^2+b^2)}{2(N+1)(N+2)(N+3)} + 2\left(\frac{N+3}{2(N+1)(N+2)}\right)^2 ab\right) \tag{3.2.T.8}$$

$$VAR_n = s_n^2 - SQ_n$$

$$VAR_E \approx VAR_{First\ Leg} + VAR_{middle,c-1} + VAR_{middle,c-2} + \cdots + VAR_{middle,1} + VAR_{Last\ Leg}$$
$$= \frac{2^{-1-c} b^2 (1 + 2^c + 2^{1+c} c)}{1 + 3c + 2c^2}$$
$$- \frac{2^{-4-2c} b^2 (1 + (1+2c)\text{Hypergeometric2F1}[1, -1-c, \frac{1}{2}, -1])^2}{(1 + 3c + 2c^2)^2}$$
$$+ \sum_{n=1}^{c-1} (s_n^2 - SQ_n) + \frac{7b^2}{24} \tag{3.2.T.9}$$

Where $c_3 = 0.525; c_4 = 0.372$;

The expectation of service time second moment:

$$E(S_E{}^2) = SQ_E = VAR_E + (E(S_E))^2$$

Average riding time on road with regression correction:

For the first customer in one loop: $Expectation(Time\ on\ Road)_1 = s_{First\ Leg}$

For the second customer in one loop: $Expectation(Time\ on\ Road)_2 = s_{First\ Leg} + s_{middle,c-1}$

……

For the last customer in one loop: $Expectation(Time\ on\ Road)_c = s_{First\ Leg} + s_{middle,c-1} + s_{middle,c-2} + \cdots + s_{middle,1}$

Therefore, the average riding time on road is

$$W_{Riding} = \frac{\sum_{i=1}^{c} Expectation(Time\ on\ Road)_i}{c} = s_{First\ Leg} + \sum_{i=1}^{c-1} \frac{i}{c} s_{middle,i}$$

$$= \frac{2^{-2-c} b(1 + (1+2c)\text{Hypergeometric2F1}[1, -1-c, \frac{1}{2}, -1])}{1 + 3c + 2c^2} + \sum_{i=1}^{c-1} \frac{(c_1 + c_2 i) i (i+3) b}{c(i+1)(i+2)} \qquad (3.2.T.10)$$

The upper bounds and approximations of average waiting time until boarding and average waiting time until delivery with regression is exactly the same as those bounds and approximations obtained for the case without regression, except for the different expression of $E(S_E)$, $E(S_E{}^2)$, VAR_E and W_{Riding}.

3.3 Simulation and Comparisons for the General-Capacity, Multi-Vehicle LMP

To assess the validity of the expressions developed in Chapter 3.2, a simple simulation of a General-Capacity, Multi-Vehicle LMTS was carried out with a program written in java. We consider a square service district with geometry $a/v_x = b/v_y = 2.5\ min = 150\ sec$, headway between train arrivals of $h = 10\ min = 600\ sec$, vehicle capacity $c = 3$, 5 or 9 and customer arrivals with batch size described by a Poisson distribution with $\lambda = 40$, 80 and 120. These parameters were selected so that the system would make sense physically. Near-optimal vehicle tours were generated by using a Traveling Salesman algorithm. Specifically, the simulation generated sets of c points, randomly and independently distributed in the square according to a uniform distribution, and a Traveling Salesman tour through these points was drawn through an algorithm that is known to generate near-optimal solutions. The algorithm implements a tour-improvement heuristic that begins with an initial solution and then improves that solution through arc exchanges ("2-exchange" heuristic) and through changes in the sequencing of the nodes in the tour ("node insertion" heuristic). More details are provided in Technical Section 3.3.T that describes the simulation experiments.

3.3.T*: Accuracy evaluation of Travelling Salesman Problem (TSP) heuristic algorithm used in the numerical experiment

In this technical section, we evaluate the algorithm accuracy using in the numerical experiment, by comparing the heuristic path length to the asymptotic Euclidean TSP lower bound.

The heuristics TSP algorithm:

1×1 square region, one point is located in the square center, and other N points are independently and uniformly distributed in the square. We use the following optimization procedures:

(1) Generate a random path;

(2) Use removals of any point j and inserting it after any point i;

(3) Improve the path locally, using replacements of sequence i, $i+1$ and $j, j+1$ with sequence i, j and $i+1, j+1$.

The asymptotic Euclidean TSP lower bound:

1×1 square region, $N+1$ points are independently and uniformly distributed in the square. David S. Johnson obtained a lower bound by computer experiment:

$0.7080\sqrt{N} + 0.522$,

where 0.522 comes from the points near square boundary which have fewer neighbors.

We obtain the following results through numerical experiment:

N random points	Heuristic Euclidean TSP	Euclidean TSP Lower Bound	Difference
2	1.286	1.748	-26.41%
3	1.657	1.938	-14.48%
4	1.944	2.105	-7.67%
5	2.177	2.256	-3.51%
6	2.374	2.395	-0.89%
7	2.543	2.525	0.74%
8	2.697	2.646	1.92%
9	2.835	2.761	2.68%
10	2.965	2.870	3.30%
11	3.083	2.975	3.64%
12	3.196	3.075	3.95%
13	3.300	3.171	4.07%
14	3.401	3.264	4.19%
15	3.498	3.354	4.28%
16	3.590	3.441	4.31%
17	3.677	3.526	4.30%
18	3.763	3.608	4.28%
19	3.844	3.688	4.22%
20	3.925	3.766	4.21%

Table 5: Comparison of TSP heuristic algorism and lower bound

The TSP length obtained by the heuristic algorithm is less than the lower bound when *N* is small, it is because one point in the heuristic case is fixed located in the center of the area, which will reduce the possible travel distance.

The TSP length obtained by the heuristic algorithm is very close to the asymptotic TSP lower bound, which is an evidence that the path provided is optimal or close to optimal, especially when *N* is no larger than 20 (we consider the vehicle capacity <=20).

Figures 19 through 23 present a sample of comparisons between the simulation results and the analytical approximations of Chapter 3.2 for the following respective cases: $c = 3, \lambda = 40$; $c = 3, \lambda = 80$; $c = 3, \lambda = 120$; $c = 5, \lambda = 80$; and $c = 9, \lambda = 120$.

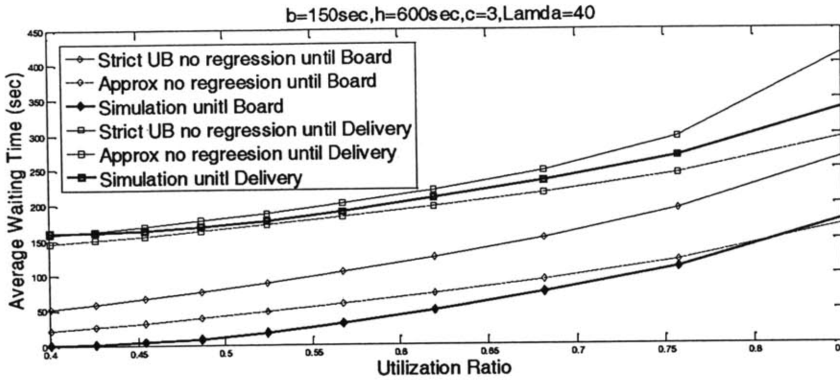


Figure 19: Simulation and analytical results when $c = 3$ and $\lambda = 40$

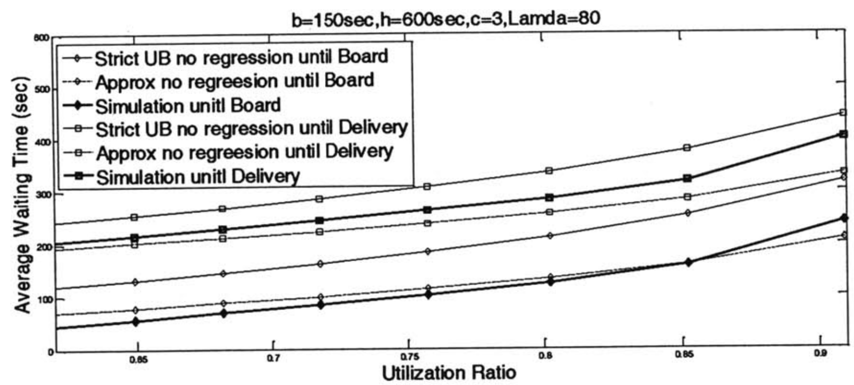


Figure 20: Simulation and analytical results when $c = 3$ and $\lambda = 80$

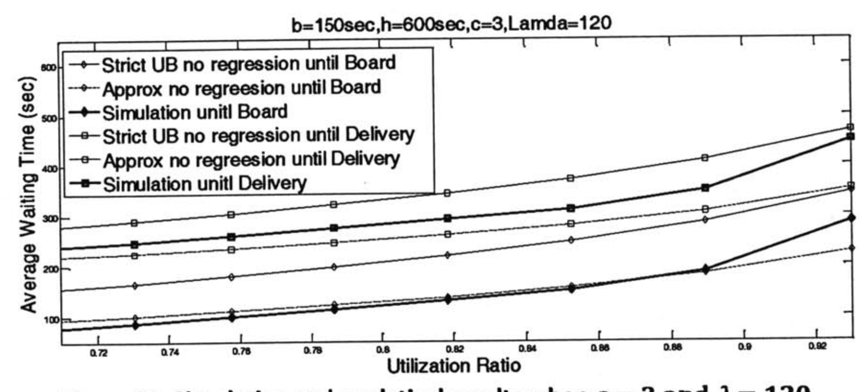


Figure 21: Simulation and analytical results when $c = 3$ and $\lambda = 120$

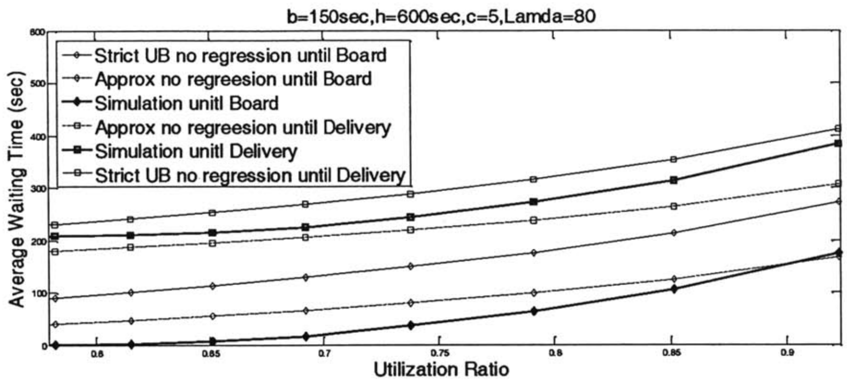


Figure 22: Simulation and analytical results when $c = 5$ and $\lambda = 80$

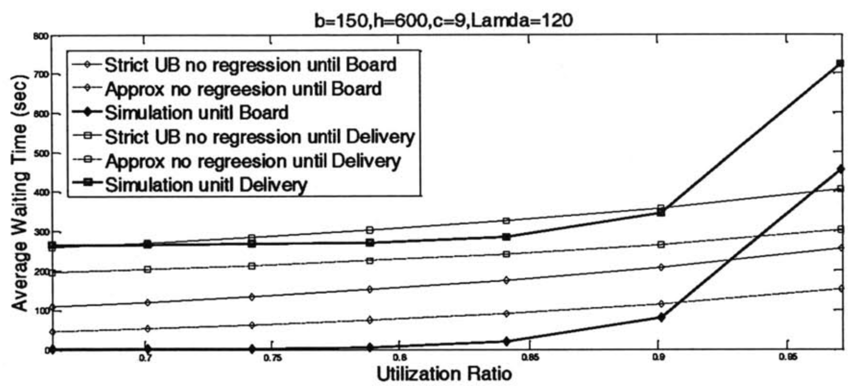


Figure 23: Simulation and analytical results when $c = 9$ and $\lambda = 120$

The horizontal axis in Figures 19-23 shows the utilization ratio $\rho = E(S_E)E(\xi_E)/mh$, while the vertical axis shows the expected waiting time until boarding a vehicle and the expected total time spent between arrival at the station and delivery at customer's destination. A comparison of the simulation results with the estimates generated through the analytical expressions of Chapter 3.2 indicated that the expressions that do *not* include a correction for the length of delivery tours (see (13) and (14)) actually perform better than the expressions that include the correction (see (15) and (16)). The explanation for this seemingly surprising observation lies in the fact that, in the absence of the correction, (13) and (14) will underestimate the expected service time (= duration of delivery tour and its second moment). This compensates for and balances out other parts of the analysis that overestimate the service time and leads to a more accurate overall approximation. Following our practice of showing only the best-performing approximations, Figures 19 – 23 therefore show only the estimates obtained through the strict cyclic upper bound (expression (17)) and the approximate cyclic upper bound (expression 18) that do not include a correction term.

When it comes to the expected waiting time until boarding a vehicle, the approximate cyclic upper bound performs very well for small vehicle size. For instance, when $c = 3$ and $c = 5$ and customer arrival intensity of 40, 80, and 120, the difference between the simulated average time until boarding and the analytical expression is of the order of 15% or less for values between 1.5 and 4 minutes, which are the most reasonable waiting time to aim for in practice. Even when the average waiting time is smaller the difference typically stays below 25%, or less than 20 seconds.

As vehicle size increases, the accuracy of the approximation of expected waiting time until boarding declines. The reason is that, when the capacity of the vehicles is large, the performance of the system becomes increasingly unstable: for example, a change of even 1 in the number of available vehicles, from some value m to $m + 1$, may result in a system transition from being nearly-saturated to being underutilized.

Turning to the estimation of expected total time until delivery, the analytical expressions work well for both small and large vehicles and for the broad range of customer arrival intensities (λ = 40, 80, and 120) examined. This can be seen in all the Figures 19 – 23. The approximation accuracy decreases somewhat as vehicle capacity gets larger, but is still good (difference less than 30% for reasonable values of total time to delivery even when $c = 9$).

4. Conclusion

This thesis has developed a set of fully analytical expressions to support the approximate estimation of the performance of a quite general version of a Last-Mile Transportation System (LMTS). Given a lengthy list of inputs about the system's characteristics (headways between arrivals of trains at the rail station, size of "batches" of customers on each train, number of vehicles in the service fleet, capacity of each vehicle, dimensions and travel-related properties of the urban district served), the expressions we have developed estimate the expected waiting time until a customer can board a vehicle, and the expected time between arrival at the rail station and delivery to the customer's destination. A number of simple simulation experiments suggest that the best-performing of the expressions we have developed approximate remarkably well the expected performance of LMTS under a broad range of conditions typical of what one may encounter in practice.

On the methodological side, the principal contribution of this research is the development of several alternative approaches for bounding and approximating the performance of a very difficult type of queuing system involving batch arrivals and requiring the simultaneous consideration of routing and queuing issues and the use of geometrical probability arguments. On the practical side, we believe that the analytical expressions we have developed can be very useful in designing LMTS, specifically in determining resource requirements for these systems, such as how many vehicles would be necessary to achieve a specified level of service and how many kilometers per day these vehicles would travel.

Future work will focus on improving the approximation accuracy for General-Capacity, Multi-Vehicle LMTS, by using a more sophisticated demand clustering and partitioning strategy and by expanding the range of the simulation inputs so that a broader range of conditions can be observed. A second area is to develop a simple set of unified guidelines for LMTS design and operation and apply these guidelines to the planning of a small actual experimental system, possibly to be implemented in a part of Singapore.

References

[1]. D. J. Bertsimas, G. van Ryzin, A stochastic and dynamic vehicle routing problem in the Euclidean Plane, Operation Research 39 (1991) 601-615.

[2]. D. J. Bertsimas, G. van Ryzin, Stochastic and dynamic vehicle routing in the Euclidean plane with multiple capacitated vehicles, Operation Research 41 (1993) 60-76.

[3]. M.R. Swihart, J.D. Papastavrou, A stochastic and dynamic model for the single-vehicle pick-up and delivery problem, European Journal of Operation Research 114 (1999) 447-464.

[4]. R. Baldacci, V. Maniezzo, A. Mingozzi, An exact method for the car pooling problem based on Lagrangean column generation, Operation Research Vol. 52, No.3 (2004) 422-429.

[5]. J-F Cordeau, G. Laporte, The dial-a-ride problem: models and algorithms, Ann Operation Reserch (2007) 153:29-56.

[6]. D.P. Bertsekas, J.N. Tsitsiklis, Introduction to Probability, Massachusetts Institute of Technology, 2008.

[7]. R.C. Larson, A.R. Odoni, Urban Operation Research, Prentice-Haill, Englewood Cliffs, NJ, 1981.

[8]. W. Kramer, M. Langenbach-Belz, Approximate Formulae for the Delay in the Queueing, System GI/G/1, 8th ITC, Melbourne, 1986.

[9]. D. Gross, J. Shortle, J. Thompson, and C. Harris, Fundamentals of Queueing Theory, 4th edition, Wiley Series in Probability and Statistics, 2008.

[10]. H. Akimaru, K. Kawashima, Teletraffic: Theory and Applications, 2nd Edition, Springer, 1993.

[11]. G. Choudhury, Analysis of the M(x)/G/1 queueing system with vacation times, The Indian Journal of Statistics 2002, Volume 64, Series B, Pt.1, 37-49

[12]. W. Whitt, The Queueing Network, The Bell System Technical Journal, Vol.62, No.9, November 1983, 2807.

[13]. O.J. Boxma, etc, Approximations of the Mean Waiting Time in an M/G/s Queueing System, Operation Research, Vol.27, No.6. November-December, 1979, 1115-1127.

[14]. T. Kimura, Diffusion Approximation for an M/G/m Queue, Operation Research, Vol.31, No.2, March-April, 1983, 304-321.

[15]. J-J Jaw, A.R. Odoni, H.N. Psaraftis, N.H. Wilson, A Heuristic Algorithm for the Multi-Vehicle Advance Request Dial-a-Ride Problem with Time Windows, Transportation Research part B, Vol.20B, No.3, 1986, 243-257.