Parcel Transportation Services: Performance Evaluation and Improvement using Markov Models

Lin Yuheng

(B.Eng., University of Science and Technology of China)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MECHANICAL ENGINEERING

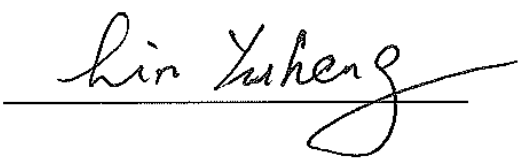
NATIONAL UNIVERSITY OF SINGAPORE

2012

Declaration

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.



Lin Yuheng

25 May 2012

Acknowledgements

A Ph.D. project cannot be accomplished singlehandedly, therefore, I would like to express my deepest appreciation and gratitude to the people who were involved in this thesis and who have shared part of my life as I embarked upon my PhD experience. I have learned as much from each of you as I have from all the courses I have taken and the books I have read.

First and foremost, I am profoundly grateful for the support and encouragement of my supervisor, Professor V. Subramaniam. I would like to thank him for sharing his knowledge and experience with me. He gave me the freedom to explore research problems using various methods. His critical remarks kept this research on course and his wealth of experience has improved the clarity of this thesis.

I would like to thank the National University of Singapore for providing me with research scholarships, research facilities, and valuable courses. I would also like to thank the wonderful and caring faculty and staff in the department of Mechanical Engineering. In particular, my deepest appreciation goes to Professor Chew Yong Tian and Lai Man On for providing me with advice and encouragement.

I would also like to extend my thanks to my colleagues in the research group, Yang Rongling, Chen Ruifeng, Cao Yongxin, Chanaka Dilhan Senanayake, and Ganish Kumar. We have worked closely together as we have discussed projects and generated new ideas. My gratitude is also extended to my friends in the laboratory of Control and Mechatronics, Chao Shuzhe, Chen Yang, Feng Xiaobing, Huang Weiwei, Wan Jie, Wen Yulin, Zhao Guoyong, Zhou Longjiang, Zhu Kunpeng, and many others, for enlightening discussions and suggestions. I am grateful for other friends I have made during my time in Singapore, Guo Guoji, Tao Fei and Wang Yueying, for your valuable friendships.

I owe my deepest thanks to my family for their unconditional love and support. Last but not least, I would like to thank my girlfriend, Fang Fang, who has lifted my spirits on the days when I have felt down.

Table of Contents

Summary

Parcel transportation services refer to the movement of small packages from or to customers. Due to dynamically changing demands and complex routings of trips, it is difficult to accurately predict performance measures on parcel transportation services such as travelling cost and service level, which is the percentage of orders that can be met within a prescribed period. There are few systematic methods published to evaluate and predict the performance of these services. Effective management tools used to determine the service price, the quantity of facilities, the range of the services, and the acceptance rules of customer demands are limited. This thesis proposes Markov models to estimate performance measures and applies optimization algorithms in order to make management decisions and improve performance of parcel transportation services.

In this thesis, parcel transportation services are characterized as Markov models based on the assumption that the vehicle travel time between customers is approximated by a hypo-exponentially distributed random variable. Two interrelated Markov processes are used to estimate transportation costs, service levels, and other performance measures. The Markov processes can be extended to resolve further related problems, such as the capacity issue, the multiple vehicles issue, the dynamic pickup or delivery issue, and services with different routing strategies. Experimental results demonstrate that the proposed Markov models are effective mathematical tools that analyze parcel transportation services and the extended problems. They are capable of providing fast and reliable estimations of various performance measures.

The proposed Markov models are able to benefit service providers in making management decisions in real-life situations. This thesis analyzes a pricing problem in order to help determine the best price for parcel transportation services. This thesis also examines an order acceptance problem in order to determine a rule for rejecting orders which are difficult to accomplish. This thesis proposes a way of designing a transportation network for the distribution center, warehouses and customers by deciding the minimum number of warehouses required, their locations, and the assignment of customers to warehouses. The proposed Markov models are able to provide reliable estimations in regards to the objective function values of these problems. Based on these estimations, satisfactory solutions can be obtained by using optimization algorithms. Therefore, the proposed Markov models in this thesis can assist transportation service providers to optimize their management decisions.

List of Tables

List of Figures

1. Introduction

1.1. Description of Parcel Transportation Services

Transportation is an integral element in numerous manufacturing and service-oriented companies. It allows businesses access to products and goods necessary to run operations. It is a critical component in collecting materials from suppliers at different locations, distributing products to customers and retailers, and transporting parts among loading docks, storage shelves and processing machines within a plant or a warehouse. Some companies manage their own transportation channels, while others outsource them to external logistics companies. Outsourcing allows companies to effectively schedule delivery vehicles and optimize transportation networks. Due to globalization, optimizing transportation for logistics companies will become more attractive in the future. Large quantities of freights and complicated transportation networks are challenging in managing logistics.

Parcel transportation services refer to the collection or delivery of small packages from or to customers. Traditionally, the transportation starts from a departure point and ends at a destination point. For example, according to a customer's request, a vehicle departs from a warehouse, travels to the customer's locations, picks up or delivers the required goods and returns to the warehouse. This kind of transportation is commonly named as spoke-hub transportation structure (as shown in Fig 1.1(a)). However, when small packages need to be transported, it is uneconomical to schedule a single trip only for one package at a time. In this case, parcel transportation services are more suitable, since it allows the service providers to fully use the capacity of a vehicle and efficiently consolidate transportation tasks for several customers. The pattern of the parcel transportation services is illustrated in Fig 1.1(b). A carrier loads parcels from the warehouse, delivers them to customers one after another in a single trip, and returns to the warehouse once all tasks scheduled for the trip are completed.

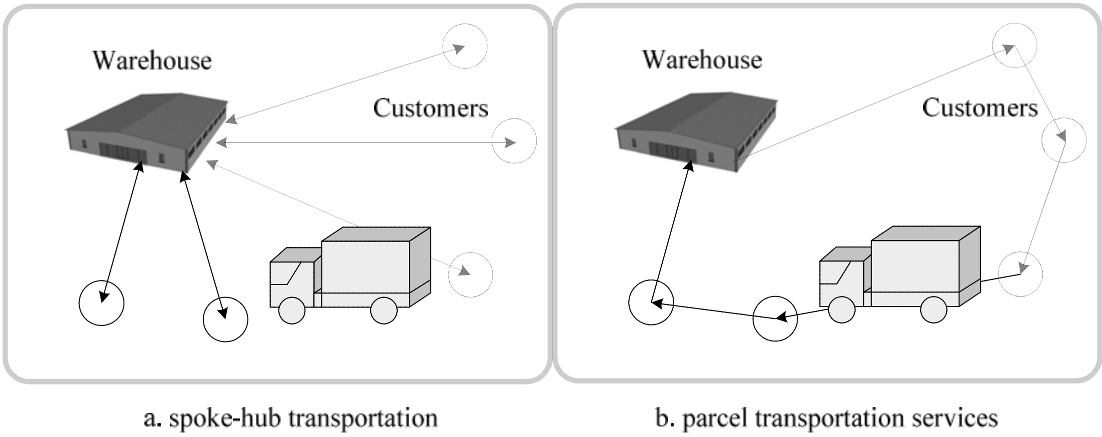


Fig. 1.1. Parcel transportation and traditional transportation

Parcel transportation services are commonly targeted at individuals, since the goods transported are relatively small. The postal systems (e.g. United States Postal Service) and third party logistics express services (e.g. DHL and FedEx) are examples of parcel transportation services. Furthermore, these services are extremely popular in our daily lives. For example, supermarkets (e.g. Fair Price and Carrefour) and electronic product manufacturers (e.g. HP and Dell) offer home delivery for food, beverages, electronic products and daily commodities. Cotton Care, a famous laundry services provider in Singapore, provides laundry pickup and delivery services to customers' doorsteps. Garbageman.com provides trash removal and cleaning services on call in South Florida. All services mentioned above can be categorized as parcel transportation services.

Business models across the world have changed drastically due to the development of the Internet, IT technologies, and electronic commerce (e-commerce). E-commerce refers to the online process of developing, marketing, selling, delivering, servicing and paying for products and services. It has led to changes in the role of logistics management and parcel transportation services. Logistics companies need to adopt suitable delivery practices in order to meet the growing demands and expectations of customers. Rather than an 'in-store service', customers prefer ordering products online, and require them to be delivered at a specific time. Therefore, e-commerce providers must effectively schedule the on-time delivery of their products. Efficient delivery of goods within a reasonably compressed period is one of the most challenging tasks for an e-commerce business. Henry Bruce, Optum's (www.optum.com) vice president of corporate marketing, claimed that "most e-commerce companies are failing in the physical delivery of products -- they have not really thought out their fulfillment strategies" (Bruce 1999).

E-commerce has three characteristics that differentiate it from traditional retailers.

- Firstly, e-commerce has an additional cost in vehicle scheduling and travelling for on-time home deliveries. This cost varies largely, and depends on patterns of transportation and methods of vehicle scheduling.
- Secondly, each product ordered from the e-commerce store is relatively small. However, the total quantity of orders is huge. For instance, Tesco.com regularly has 750,000 customers and 200,000 orders per week, but one order may only worth several dollars (Asdemir et al. 2009).
- Thirdly, e-commerce orders are relatively dynamic. Customers' orders appear randomly, so it is difficult to predict the ordering time as well as the quantity of orders. Moreover, the required response time and delivery time are relatively short. Therefore, it is necessary to develop a vehicle-dispatching framework that handles dynamic demands of consumers in a short time frame.

Parcel transportation services have the potential to meet the requirements of e-commerce, which need to handle a cost effective shipping operation to a huge number of customers in a dynamic environment. Parcel transportation services that are tactically managed are capable of providing a relatively efficient solution to transportation problems in the *last mile*[1] of the supply chain.

1.2. Motivation

The competition is extremely high in parcel transportation services. Companies make relentless efforts to reduce cost and improve service quality to beat their competitions. In order to survive this fierce competition, service providers must improve their cost and revenue management technologies.

Transportation service providers devote most of their efforts in reducing costs, since cost plays a critical role in transportation services. In order to reduce costs, various technologies are applied to address how to collect and deliver parcels from and to various locations. Numerous vehicle scheduling and route planning strategies have been proposed by researchers.

Revenue management is the application of disciplined analytics that predict customer behavior and optimize product availability and price, to maximize revenue and profit. Minimal research has focused on developing models for the revenue management of parcel transportation services. Applicable analytical models may be derived from prior research on product revenue and manufacturing operation managements. These methods have to predict the demand and apply pricing approach, which is widely applied to increase product sales (Dong et al. 2009), and order acceptance rules, which is widely

[1] *Last mile* is a term used in supply chain management and transportation planning to describe the movement of people and goods from a transport hub to a final destination.

utilized in job shop environment (Ebben et al. 2005). However, research works in the literature do not fully take into consideration of a comprehensive measure including revenue, cost and future impact of the decision in revenue and cost management, since it is too difficult to analyze the vehicle routing and scheduling issue in the process of optimizing service price or order acceptance rules. Therefore, estimating performance measures and the objective of revenue management tends to be incomplete and biased.

In order to properly apply revenue management tools for parcel transportation services, it is necessary to decide a comprehensive objective including key performance measures, such as revenue reflecting the customer demands, cost of vehicle travelling and quality of service. Various performance measures are required to be estimated using a systematic method based on different strategies of vehicle routing and scheduling in a dynamic environment. There is little literature that specifically addresses a systematic method which is able to predict most of the performance measures for parcel transportation services in dynamic vehicle routing conditions. In this thesis, the author will discuss the procedure for building a systematic method, which can accurately estimate various performance measures in a short time frame to different situations. The thesis also elaborates the use of management decision tools that can be utilized by transportation service providers to manage their business effectively and thereby help them to survive in the global market.

1.2.1. Performance Measures

Performance measures reflect the state of the business, and are commonly used by companies to evaluate the success of their business towards certain goals. Without the ability to properly measure performance, companies are in no position to analyze their business and improve their efficiencies. In parcel transportation services, a variety of performance measures are used. Some commonly used performance measures in practice are highlighted as follows.

a. Travel distance and transportation costs

Travel distance and transportation costs are two important performance measures in parcel transportation services. Transportation costs include the fuel expenses, which are proportional to the vehicle travel distance, the expenses on the usage of vehicles, and the labor costs of drivers. In general, transportation costs account for a high proportion of the national expenditures in North American and European countries (Crainic and Laporte 1997, Larsen 2000). For example, the road transportation costs were about 5% of the United States' gross domestic product (GDP) in the past 10 years. According to CSCMP's 22nd annual state of logistics report (Rosalyn, 2011), transportation costs in the United States reached $768 billion USD, of which 78% were from road transport. The Logistics and Supply Chain Management Key Performance Indicators Analysis of Canada (2006) showed that total transportation costs were about 2.5% ~ 10% of the product sales revenue in Canada. Another survey shows that one third of customers agree that transportation costs significantly affect their purchase decisions (Reynolds, 2001). Therefore, transportation costs are huge and they significantly influence the product sales and global economy.

Furthermore, liquid fossil fuels are the main energy sources for transportation. Transportation consumes more than 60% of the oil supply of the world. The "Repowering Transport Project White Paper" (2011) predicted that the energy consumption by transportation will continue to grow and will be 40% higher than current levels by 2030. Significant energy consumption and CO_2 and other vehicles emissions have attracted the government's attention. All of the aforementioned concerns related to transportation costs, energy consumption, and pollution have prompted us to develop more efficient methods to reduce the distances that vehicles travel.

Researches on parcel transportation services studied optimization of total travel distance or total transportation costs (Bräysy and Dullaert, 2003; Polacek et al., 2004; Mester and Bräysy, 2005). However, the optimization objectives of such studies may not be suitable for a dynamic situation, in which service providers receive new orders at any time and the service may be endless (Psaraftis, 1995). Therefore, performance measures for such situations are better evaluated using average values, such as average travel distance for each customer and average transportation costs per unit time.

b. Revenue

Revenue is defined as funds received by a company from the sale of products or services, and it depends on customer demand. Hence, forecasting customer demand is crucial. Since demand for parcel transportation services are stochastic and dynamic, it is difficult to predict in advance which customers will be willing to pay for services and how many tasks can be completed within a specified operating period. Proper probability distributions or stochastic models can be helpful in forecasting customer demand. Average revenue is proportional to the average number of customers fulfilled per unit time, and indicates the customer demand rate and the company's ability to handle customer requests.

c. Vehicle utilization

Vehicle utilization is defined as the percentage of time that a vehicle is engaged in providing transportation services. Vehicles continue to travel between customers and the warehouse if there are pending customer requests; otherwise, they are idle while waiting for new orders. When demand is high, vehicle utilization increases accordingly. To meet additional demand, managers may have to increase the number of vehicles in use. In contrast, if vehicle utilization is low, the potential exists to serve additional customers. In such a situation, managers may need to reduce the number of vehicles or increase demand.

d. Average number of customers waiting for services

The average number of customers waiting for services is similar to the average length of a queue. A large number of customers waiting for services may indicate service inefficiency or lack of vehicles and other resources.

e. Order-to-delivery time

Order-to-delivery time represents the time elapsed between placement of an order by a customer and delivery of the product to the customer. Order-to-delivery time reflects the response time from a transportation service provider's point of view and waiting time from a customer's point of view. Long order-to-delivery time fails to meet customer expectations, and may result in lose customers. All managers attempt to shorten this period by improving the efficiency of their transportation services.

f. On-time fulfill rate and service levels

In logistics services, on-time fulfill rate and service level are defined as the percentage of orders that can be met within a prescribed period. Usually, transportation service providers enter into agreements with the customer stating that products will be delivered to their destinations within a specific period. Delays result in customer dissatisfaction with the service provided. Too many delayed deliveries will tarnish the reputations of logistics companies and affect future profits. As a result, logistics companies need to compensate customers for unsatisfactory experiences. However, it is difficult for transportation service providers to consistently achieve on-time delivery under dynamic demand and traffic conditions. Therefore, they must endeavor to reduce the possibility of delays and provide high-quality transportation services.

A successful transportation service must take into consideration service level, which represents the quality of services that customers receive. A high service level indicates that most products are delivered on time as specified by customers. A company providing good services has a competitive advantage over its competitors (Mentzer et al., 2004). The aim of logistics is turning from minimizing distribution costs towards increasing customer service quality (Lehmusvaara, 1998). Therefore, an increasing number of companies are paying attention to efficient planning to provide quick responses to customer orders while at the same time maintaining a high service quality.

g. Profit

A comprehensive measure of performance is profit, and a key objective of all companies is to maximize both short-term and long-term profits. In logistics services, short-term profit is equal to the difference between the revenue earned from the business and transportation costs, whereas long-term profit takes into consideration service level. Failure to satisfy customer requirements will result in customer complaints, a reduction in orders and a tarnished company reputation. In this thesis, profit is selected as the overall performance measure for parcel transportation services, and is defined as the difference between the revenue and the cost associated with operating the business. The cost includes transportation costs and penalties incurred by delivering low quality services.

1.2.2. Methods for Estimating the Performance Measures

Efficient estimations of performance measures are needed, and may be obtained using evaluation tools (Fig. 1.2). Simulation is one such widely used tool. By providing information on customers, the logistics company, routing algorithms, and decision variable values, the simulation generates an estimation of performance measures. Different estimations of performance measures are obtained as the values of decision variables change. In the end, the best value for the decision variables can be determined using optimization algorithms. However, it may take time for the process using simulation to obtain accurate estimations and optimized decision values.

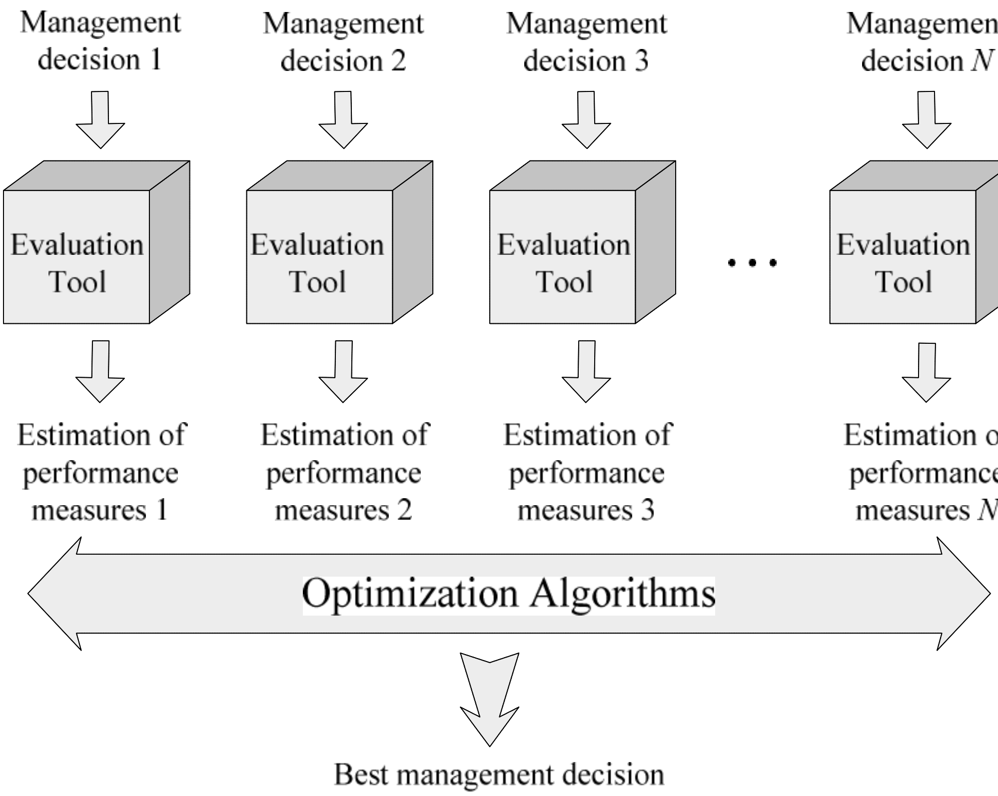


Fig. 1.2. Management decision-making structure involving estimations of performance measures using evaluation tools

Another method for estimating performance measures is to mathematically construct a function which consists of all the decision variables. This method obtains the best value for the decision variables to minimize or maximize one of the performance measures by solving a number of differential equations. One example in this category is the linear regression method (Confessore et al. 2008), which constructs a polynomial cost function to approximately fit historical data. The function obtained from this method is fast in calculating costs but has lower accuracy.

In the literature, mathematical methods based on queuing theory attempt to determine the lower bound of transportation costs, which is approximated using a function of demand rate. The Queuing theory assumes that the arrival of customer requests in parcel transportation services is a random process, which is usually a generalization of Poisson process. Customer locations are assumed to be uniformly distributed on a Euclidean plane. It is also assumed that the travel time between two successive customers in a planned vehicle trip has a mean τ and a variance σ^2 (Bertsimas and Ryzin, 1991). One may approximate these predictable values through experience or by using historical data. Since the variable of travel time depends on routing strategies, the accurate evaluation of travel time is important but complex. Based on this probability, parcel transportation behaves analogously to a queuing system with generally distributed service time. Typically, the cost function generated from the queuing theory provides an upper or a lower bound. They also further extended their work on the calculation of the upper and lower bound for transportation costs and average customer waiting time in the cases of capacitated vehicles, non-uniform spatial distributions, and general renewal processes for arrivals (Bertsimas and Ryzin, 1993a; Bertsimas and Ryzin, 1993b). However, it may be difficult to extend these methods to calculate other performance measures, such as the distribution of customer waiting time and service levels. Furthermore, since these methods only provide boundaries for transportation costs, it may not be possible to optimize the decision values based on the boundary estimations.

This thesis constructs a Markov model for parcel transportation services. The Markov model is an analytical model derived from the stochastic process of the queuing system with generally distributed service time. Therefore, the model provides a systematic view of the problem. If customer demand rate, routing algorithms and decision values are provided, this queuing-based stochastic model estimates performance measures using probabilities and transition matrices. Results show that this method is more efficient, flexible, and accurate in estimating performance measures. Based on these estimations, logistics companies may be in a better position to make management decisions.

1.2.3. Application of Performance Measures Evaluation in Parcel Transportation Services

Using estimations of performance measures, issues at the management level may be analyzed to obtain optimal solutions. For example, when a logistics company starts services in a new urban area, the manager can use estimations of performance measures to define profitable service regions and provide parcel transportation services to these areas. The manager may also use the estimations to set a suitable price for the service, operate the business with a minimum number of vehicles, warehouses, and resources. In addition, the manager may utilize the order acceptance rule to insure the business is more profitable. For logistics services that have been in operation for a long period, the manager must remain up-to-date on technology and market changes, and may evaluate again the services based on the performance in order to adjust service region, price, number and locations of warehouses, and operation rules accordingly. In these cases, the manager's decisions should optimize profit, and decision variables may involve service price, number of vehicles, number of warehouses, and job acceptance rules. Resolving such optimization issues is the main concern of this research.

An example of decision-making is related to the pricing approaches for online purchase and home delivery services. Nowadays, stores allow customers to order products from "virtual storefronts" through the Internet, and deliver ordered products to locations that customers specify for additional service charges. Delivery services are provided by an external logistics company, which needs to make careful decisions on price, effectively reduce transportation cost and maintain high quality of services. Published literature in this field seldom analyzes the pricing issue related to dynamic vehicle routing, which economically utilizes a single vehicle trip with multiple delivery stops. The goal of this thesis is to provide a stochastic model to help service providers address management decision-making issues related to a dynamic vehicle routing based environment.

1.3. Contribution of this Thesis

In this thesis, a new stochastic approach based on a Markov model has been developed to properly and accurately estimate performance measures for parcel transportation services. Compared with simulation methods, the Markov model provides fast estimations of various performance measures including transportation costs and quality of services. Furthermore, the model is flexible and can adapt to extensions for various parcel transportation services.

Based on the accurate evaluation of performance measures, parcel transportation services can be systematically investigated. Service providers can effectively manage their business and make management decisions to improve service performance. Three practical transportation problems faced by numerous logistics companies are analyzed in this thesis: the service-pricing problem, the transportation network design problem and the order acceptance problem. The new stochastic approach is able to provide fast and accurate estimations of the objective function values for these problems. Based on the estimations, optimization algorithms can easily converge to the best solution that can be utilized for management decisions.

1.4. Outline of the Thesis

The rest of this thesis is organized as follows. Section 2 provides a literature review on modeling parcel transportation services and evaluating performance measures. Although some vehicle routing strategies are discussed, the new stochastic approach of modeling parcel transportation services in dynamic circumstances is the major focus of this research. Section 3 estimates the performance measures for parcel transportation services utilizing a Markov model. The model is then verified through comparisons between the proposed model and simulations in several numerical experiments. In section 4, the modifications to the Markov model are presented for different scenarios in the parcel transportation services. In section 5, the applications of the proposed Markov models in three practical problems are investigated. Optimal solutions for these practical problems are obtained. Finally, this thesis concludes with a summary of the key findings of this research and proposals for future research.

2. Modelling of Parcel Transportation Services: State of the Art

2.1. The Problem of Vehicle Routing in Parcel Transportation Services

The management of parcel transportation services involves solving a vehicle routing problem (VRP). The VRP seeks an optimal routing schedule for a fleet of vehicles in order to efficiently serve customers scattered in a pre-defined region. The VRP is one of the most prevalent topics studied in the field of operational research (Golden et al., 2008). The VRP is stated as determining optimal routes on an undirected complete graph $G=(\mathbb{N},\mathbb{Z})$. $\mathbb{N}=\{n_0,n_1,\cdots,n_N\}$ denotes the vertex set of nodes, where n_0 is the warehouse and $n_1\cdots n_N$ are N customers which must be visited exactly once. $\mathbb{Z}=\{(n_i,n_j):n_i,n_j\in\mathbb{N},i\neq j\}$ denotes arcs between nodes with non-negative weights representing travel distances and associated travel time. The travel distances and times on arcs are usually described in a symmetrical matrix, which indicates that the distance from node *i* to node *j* is the same as that from node *j* to node *i*.

Numerous routing algorithms are proposed to provide the best vehicle routing schedules for a group of customers whose demands are known before the start of the services. Exact algorithms, such as dynamic programming, Lagrange relaxation and column generation (see Laporte, 1992), may be used to solve problems that involve a small number of customers. However, the VRP has been considered non-deterministic polynomial-time hard (NP-hard), in which the time spent on obtaining a solution increases exponentially when the total number of customers increases. A vast effort has been devoted to computing optimal solutions for NP-hard problems. However, numerous approximation methods and heuristics are proposed to effectively speed up the search for a satisfactory solution instead of exhaustively searching for the optimal solution. These methods are categorized into construction heuristics, improvement heuristics and metaheuristics (Ropke, 2005).

- Construction heuristics gradually build a feasible solution with no improvement phase. The Clarke and Wright saving algorithm (Clarke and Wright, 1964) and the sweep algorithm (Gillett and Miller, 1974) belong to this category.
- Improvement heuristics implement changes of decision values from initial solutions, which may be obtained from constructive heuristics, to generate more effective solutions. Local search algorithms (Bräysy and Dullaert, 2003; Chaovalitwongse et al., 2003; Mester and Bräysy, 2005) and neighborhood search algorithms (Bräysy, 2003; Polacek et al., 2004) belong to improvement heuristics only when they perform operations that lead to improved outcomes relative to the objective.
- Metaheuristics implement changes of decision values that lead to both better solutions and worse solutions. Metaheuristics is applied in order to prevent the process from adhering to the local optimum to find a more effective solution after further changes. Tabu search algorithms (Lau et al., 2003; Ho and Haugland, 2004), genetic algorithms (Wei, 2003), ant colony optimization algorithms (Kuo et al., 2004), and simulated annealing algorithms (Amberg et al., 2000) belong to this category.

In order to address different aspects of the service requirements, researchers have developed various extensions of the VRP in operation research. Some extensions extensively studied in the literature are highlighted as follows. The solutions of these problems can be derived from the routing strategies mentioned previously.

- Capacitated vehicle routing problem (CVRP). In the CVRP, a non-negative weight is attached to each customer and the sum of weights in any vehicle route must not exceed the vehicle capacity (Lysgaard et al., 2004). In addition to the travel distance, this problem is required to minimize the number of vehicle trips.
- Vehicle routing problem with time windows (VRPTW) adds a time window constraint to each customer's request, indicating an opening and closing time within which services can be performed. In this situation, the vehicle must visit customers within designated time windows in order to avoid service failure. This time window constraint may cause difficulty in finding suitable solutions. Taillard et al. (1997) suggested a soft time window constraint. In this condition, the vehicle is permitted to arrive before the opening time of a specific service and wait until it is allowed to start the service. However, if the vehicle arrives later than the closing time, a penalty will be applied to the transportation cost.
- Vehicle Routing Problem with Length Constraint (VRPLC). Since the maximum continuous working period of a driver is restricted in some countries, the length constraint restricts that the length of a planned vehicle route is not allowed to exceed a prescribed limit (Li et al., 1992; Nagarajan and Ravi, 2012).
- Multi depot vehicle routing problem (MDVRP). This problem is studied to manage vehicles from different warehouses while serving the same group of customers (Renaud et al., 1996; Lim and Wang, 2005). The MDVRP can be extended to a

problem which seeks the best locations for the warehouses (Chan et al., 2001).

- Vehicle routing problem with pickup and delivery (VRPPD). Two different kinds of requests are studied in this case, namely pickup and delivery requests. In delivery requests, products are required to be loaded on board and transported to specific locations. The pickup requests allow the vehicle to utilize spare capacity to collect goods from customers, and to carry the goods back to the warehouse. The VRPPD can be further classified into three categories.
  - Transportation services following a delivery-first-pickup-second rule are usually characterized as a vehicle routing problem with backhauling (VRPB) (Ganesh and Narendran, 2007; Tavakkoli-Moghaddam et al., 2006). For the convenience of truck loading, the problem specifies the first-in-last-out (FILO) rule, which requires all delivery requests be satisfied before any pickup requests are considered.
  - Services with mixed pickup and delivery are characterized as a mixed vehicle routing problem with backhauling (MVRPB) (Wade and Salhi, 2002; Sural and Bookbinder, 2003; Zhong and Cole, 2005). The MVRPB allows pickup and delivery services in any sequence on the route, due to the widespread use of side-loaded trucks. A practical example of MVRPB for a logistics company in Hong Kong is analyzed by Cheung and Hang (2003).
  - Vehicle routing problem with simultaneous pickup and delivery (VRPSPD) allows each customer to make a pickup request and a delivery request (Mitra, 2005; Dell'Amico et al., 2006; Montane and Galvao, 2006; Bianchessi and Righini, 2007; Gribkovskaia et al., 2007).

2.2. The Problem of Dynamic Vehicle Routing in Parcel Transportation Services

The dynamic vehicle routing problem (DVRP) addresses concerns regarding uncertain demand and dynamic traffic conditions in parcel transportation services. Most real life transportation scenarios operate under dynamically changing information and unpredictable circumstances. Customer demands are stochastic and dynamic, and it is impossible to predict when a customer will place an order and where goods will need to be picked up or delivered. In addition, traffic conditions change over time. The vehicles may occasionally have accidents or experience delays. The uncertainty of the demand and the dynamism of traffic conditions make the problem much more complicated.

In order to look into the transportation problems from a dynamic perspective, logistics providers resort to advanced communication and information technologies. The Geographical Information System (GIS) and the Global Positioning System (GPS) provide location maps and exact vehicle positions, and therefore have become an integral component of vehicle routing operations (Ghiani et al. 2003). In addition, radio frequency identification (RFID) technology enables the tracking of products and spaces in vehicles throughout the entire transportation process. These technologies utilize real-time data and enable dynamic transportation services within hours of a request being made.

2.2.1. Comparison between VRP and DVRP

Compared to the static VRP, the DVRP has several distinct features.

- In the DVRP, the time dimension is essential, as circumstances in a dynamic problem change frequently over time. Planning for services depends on the arrival of new orders and variations in external conditions. The current plan may not be valid a few minutes later.

- In the DVRP, it is impossible to obtain the accurate information in advance. Probabilistic information may be summarized from the statistical analysis on previous data (Psaraftis, 1988).
- In the DVRP, high-speed computations are required. In static settings, the dispatcher may be able to afford the luxury of waiting for a few hours in order to get a high quality or optimal solution. However, in a dynamic setting, the dispatcher requires a feasible solution to the current problem within a limited time frame (Psaraftis, 1995).
- In most cases, a DVRP can be treated as a queuing problem, in which customers wait for services while vehicles work as servers. If the rate of customer demand exceeds a threshold, the system will become congested (Larsen, 2000).
- Last but not the least, the objective function may be different. Traditional static objectives such as minimizing the total distance travelled may not be appropriate in a dynamic setting. In dynamic circumstances, it is difficult to determine the number of requests accomplished in a specific period of time, and a logistics company will provide services for a long time. Therefore, the objective function is better represented in average values, such as the average cost for each customer or the average profit per unit time. Additionally, the service level should be an objective considered while solving the DVRP. The service level reflects the quality of service received by customers, and affects customers' options in the future.

2.2.2. Routing Strategies for Dynamic Vehicle Routing Problems

The DVRP addresses the key concerns of parcel transportation service providers, which are to manage uncertain demands and dynamic traffic conditions by proper planning and scheduling of the transportation system. In the Stochastic Vehicle Routing Problem (SVRP), it is assumed that customer demand follows certain probability distributions (Secomandi, 2001). The problem is defined on a graph of N fixed nodes, which represents the locations of customers. Each customer will require a visit only with a certain probability. However, in a DVRP, not only is the demand uncertain, but the number and the locations of customers are also unknown. Researchers usually assume that the customer demands appear according to a Poisson process and customers' locations are independently and uniformly distributed in the service region. (Bertsimas and Ryzin, 1991).

Numerous routing strategies are proposed, and most of them address the issue of dynamic customer demands. Two steps are usually involved in these routing strategies (Tighe et al., 2004; Potvin et al., 2006). The first step is to generate an initial plan for known requests. In the first step, the problem is similar to a static one, and all the methods for the static VRP can be used to generate initial solutions for the DVRP (Psaraftis, 1988). More discussions focus on the second step of the online routing strategy for the DVRP, which provides a rule for the adjustment of the existing routing schedule when a new service request appears. One simple and quick strategy is called First-Come-First-Serve (FCFS), which allows the new service request to be processed at the end of the existing routing schedule. A number of researchers prefer insertion algorithms, which seek the best insertion place in the existing routing schedule for the new requests. Zhu and Ong (2000) claimed that insertion algorithms generate better solutions and quicker responses to new requests. Other researchers resort to metaheuristics to refine the routing schedule after insertion. For example, Pankratz (2005) generated a solution pool by randomly swapping and re-arranging the delivery sequence in order to find a better solution using a genetic algorithm. The routing plans are rescheduled through complicated methods whenever a new request appears or the traffic condition changes. Ferrucci et al. (2013) and Tirado et al. (2013) used Tabu search to refine and update the solutions. They took the potential effect of future orders into consideration in the current routing plan. These approaches require considerable computational efforts, and various routing strategies and dynamic vehicle schedules complicate the estimation of performance measures for parcel transportation services.

2.2.3. Comparison Strategies for Dynamic Vehicle Routing Problems

Numerous routing strategies were proposed in past research. Since most strategies are developed based on randomly generated cases, determining the best routing strategy is difficult. In order to compare different online routing strategies and evaluate performance, Sleator and Tarjan (1985) proposed a competitive analysis method, which has been widely used for scheduling and financial decision making (Manasse et al., 1990; El-Yaniv et al., 1992). The competitive ratio is measured by the worst case ratio between the objective value gained from the online strategy for a sequence of randomly generated requests and the optimal value gained from an algorithm which knows the entire sequence in advance. However, it is difficult to achieve the optimal solution in most cases. An offline solution obtained from algorithms with all information known in advance is usually used as a benchmark to replace the optimal solution. The purpose of the competitive analysis is to find out the largest gap between the online strategy and the benchmark. Ausiello et al. (1994) first suggested competitive analysis for the DVRP.

The objective of this research is to minimize the completion time until a certain group of the requests are served. Special cases have been used to prove that no online strategies are able to achieve a competitive ratio lower than two. Other researchers improved the competitive ratio by revising the objective function of the problem or by restricting the powers of the offline routing algorithms in order to more closely resemble the online strategy. For example, Ausiello et al. (1995, 2001) discussed the case where the objective was to minimize the time taken for a vehicle to serve a certain group of customers and return to the warehouse. Blom et al. (2000) and Krumke et al. (2002) suggested applying a fair adversary to calculate the competitive benchmark. The fair adversary, in this case, refers to imposing a restriction that the vehicle in the offline case can only move in the direction where pending requests are present. The competitive analysis provides a way to compare the performance of various online strategies for the DVRP. However, this method only provides a comparison of the worst-case scenarios instead of evaluating the general case scenarios of each routing strategy.

Other research focuses on computation of an upper or lower bound. Bertsimas and Ryzin (1991) revealed that parcel transportation services are similar to queuing systems. In that research, five routing policies have been proposed for the problem, which are the stochastic queue median, partitioning, travelling salesman, space filling curve and nearest neighbor policies. The results for queuing systems (see Kleinrock 1976) are used to calculate the upper and lower bounds of the expected customer waiting time in either heavy traffic (when the arrival rate of demands is high) or light traffic (when the arrival rate of demands is low) situation. Bertsimas and Ryzin concluded that the stochastic queue median policy yielded the best result in a light traffic situation but was not stable in a heavy traffic situation. Papastavrou (1996) developed a new routing policy for the DVRP. The lower bound was calculated, and numerical results showed that the policy performed well in both light traffic and heavy traffic situations. Ghiani et al. (2007) calculated the lower bounds for two deferment policies and an insertion policy based on the formula proposed by Bertsimas and Ryzin, and concluded that the insertion policy outperformed the others. Bertsimas and Ryzin (1993a,b) also calculated the bounds for transportation costs and average customer waiting time in the cases of non-uniform spatial demand distributions, and general renewal processes for customer arrivals. Bullo et al. (2011) concluded that a uniform spatial density of demand leads to the worst possible performance of customer waiting time, and the deviation from uniformity in the demand distribution will strictly lower the optimal expected waiting time. They also claimed that providing higher priority of service to certain demands would result in a reduction of optimal expected waiting time for non-uniform density demand distributions. These studies involved a large amount of effort in order to approximate and calculate the lower and upper bounds. However, the results only reflect the performance of routing algorithms in special cases.

2.3. Evaluation of Performance Measures

The study of parcel transportation in dynamic conditions is interesting and is not limited to the design of routing strategies. An analysis of different routing algorithms reveals that the differences between results obtained by various algorithms are small. This means that improving vehicle routes using different routing algorithms is trivial. Moreover, different algorithms have specific advantages in different cases. A simple routing algorithm may be more economical and efficient than other sophisticated algorithms in some cases. Spending effort to seek the best algorithm is not always cost-effective. A quick and accurate estimation of the performance measures may be more meaningful than a detailed vehicle routing schedule in the management of transportation services (Bruns et al. 2000; Wasner and Zäpfel 2004).

Simulation is one of the most popular methods used in estimating the performance measures. Past research identifies several simulation structures (Du et al., 2005; Hanshar and Ombuki-Berman, 2007; Barbucha and Jedrzejowicz, 2008; Xiang et al., 2008), as summarized in Fig. 2.1.

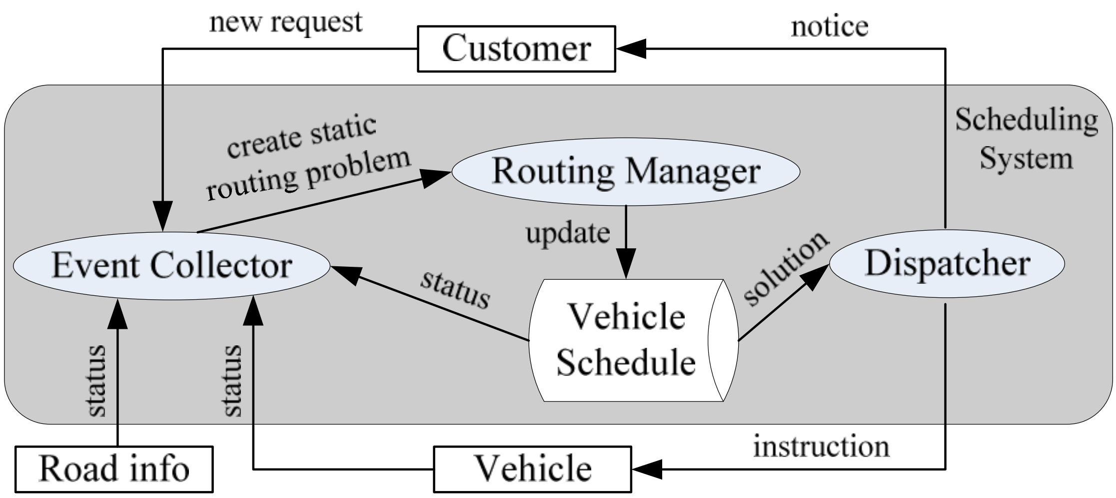


Fig. 2.1. The framework of simulating the Dynamic Vehicle Routing Problem

- The customer module generates customer requests with arrival time, locations, and due time based on certain probability distributions.
- The scheduling system processes customer requests and generates vehicle schedules.
- The vehicles follow instructions from the scheduling system, and continuously update the scheduling system regarding their statuses. For example, a status could include whether the vehicle is idle, whether the vehicle is experiencing a

breakdown on the road, and the location of the vehicle itself.

- The road module indicates the route network between customer locations and the warehouse. It calculates the distances or travel time, and updates the scheduling system about traffic conditions.

Within the scheduling system, there are four components, which are the event collector, routing manager, vehicle schedule, and dispatcher.

- The event collector gathers new requests, road information, the vehicle status with its current schedule, when customers make new service requests or the vehicle status changes. It creates a static VRP with all the information and passes it to the routing manager.
- The routing manager resolves this static VRP based on certain routing strategies and updates the vehicle schedule with the current optimal solution. In past research, these routing strategies are usually described by pseudo-codes (de Oliveira et al. 2008; Jun et al. 2008; Angelelli et al. 2009) or program structure diagrams (Fleischmann et al. 2004; Ahmmed et al. 2008; Xiang et al. 2008). Within the resolution period, the scheduling system is locked, which means that new customer requests, updated traffic conditions or vehicle status changes will only be handled after the routing manager has resolved the current VRP and unlocked the scheduling system.
- The dispatcher retrieves the solution from the vehicle schedule, and informs the customer whether a request is accepted or rejected, as well as an estimated time in which the service will be provided to the accepted request. It also instructs drivers to follow the current schedule.

After a large number of requests are generated, estimations of performance measures are obtained based on data collected from simulations. Simulation is easily implemented, and is flexible for various scenarios. However, it may take a long time for the simulation to obtain a reliable estimation.

Another method to evaluate performance measures is based on mathematical formulas. Mathematical programming is commonly used to formulate the VRP (see Laporte 1992). Using mathematical programming, the minimum distances or costs may be obtained. Based on the travel distance and cost provided by mathematical programming of a VRP, an algorithm can be used to determine the frequency of the vehicle travelling in an inventory management problem (Rajeshkumar and RameshBabu, 2006) and the location of depots in the network design of transportation services (Wasner and Zäpfel, 2004). However, it is complicated to represent the DVRP in a mathematical programming formula, since the situation changes over time. Haghani and Jung (2005) tried to use a mathematical programming formula to analyze the upper and lower bounds of the travel distance in a DVRP. They reported numeric results of only 10 demands in several discrete time intervals. Therefore, this approach is not efficient to evaluate the performance measures of parcel transportation services.

A few researchers have attempted to calculate performance measures based on a mathematical function of specific random variables. No programming codes or graphs are used to represent routing strategies in this research stream. The effects of the routing algorithms on the final performance measures are represented by parameters in mathematical functions. For instance, Confessore et al. (2008) tried to construct a travelling cost function of the average span of customer time windows based on experience and historical data. This historical data is generated from certain routing strategies. Linear regression is used to calculate the value of parameters in the cost function. However, the cost function in this research is simple and lacking of mathematical deduction and proof. Hence, performance measures estimated by this function are not sufficiently accurate.

Continuous approximation models are popular estimation methods for determining vehicle travel distances. The key feature of these models is that the total travel distance d is only estimated by a function of the area A of the service region R and the spatial density $\delta(x, y)$ of the customer locations (x, y) (Langevin and Mbaraga, 1996). The average Euclidean distance between the warehouse and a customer within the region is calculated as follows.

$$\bar{l} = \frac{\iint_R l_0(x, y)\delta(x, y)dxdy}{\iint_R \delta(x, y)dxdy} \tag{2.1}$$

$l_0(x, y)$ is the distance from the warehouse to location (x, y). If customers' locations are uniformly distributed in the region, the average distance from the warehouse to any customer's location is simplified as follows.

$$\bar{l} = \alpha_1\sqrt{A} \tag{2.2}$$

α_1 is a constant which depends on the shape of the region and the location of the warehouse. Beardwood et al. (1959) proved the optimal total travel distance between N customers uniformly distributed in the service region as Equation (2.3).

$$d = \alpha_2\sqrt{AN} \qquad N \to \infty \tag{2.3}$$

In this equation, α_2 is a constant parameter which depends on the shape of the region. A number of researchers have tried to estimate the parameters α_1 and α_2 in differently shaped regions (refer to Langevin and Mbaraga, 1996).

Continuous approximation models have wide applications in the study of transportation problems. In warehouse location and relocation problems, Bruns et al (2000) divided the travel distance into several running distances and two stem distances. A stem distance is the average distance from the warehouse to the first customer in a trip or from the last customer to the warehouse, which is obtained by Equation (2.2). A running distance is the average distance between two consecutive customers, and is proportional to the square root of the area of the service region, which is derived from Equation (2.3). Newell and Daganzo (1986) tried to partition a region of delivery services into a number of sub-regions and decide the size of each sub-region. In that research, the objective was to minimize the total travel distance which was estimated by continuous approximation models. Daganzo and Newell (1985) studied the frequency of product deliveries by exploring the trade-off between the increasing inventory cost due to accumulation of products and the decreasing travelling cost due to longer and more efficient vehicle routes. The travelling cost was also evaluated by continuous approximation models. Geunes et al. (2007) analyzed a pricing problem in delivery services and derived the cost function based on continuous approximation models proposed by Haimovich and Rinnooy Kan (1985).

Bertsimas and Ryzin (1991), Papastavrou (1996) and Ghiani et al (2007) assumed that customer inter-arrival time followed an exponential distribution and customers were randomly dispatched over the map in a uniformly distributed pattern. The travel time of the vehicle was a random variable with a certain probability distribution, where the mean and variance can be estimated by continuous approximation models. Bertsimas and Ryzin (1993a, 1993b) studied the VRP with multiple capacitated vehicles, non-uniform spatial distributions and general renewal processes for arrivals with these models. These research extended results by Christofides and Eilon (1969) and the queuing theory in order to estimate the upper bound and lower bounds of average customer waiting time. However, the bound analysis of average waiting time was not enough to measure the overall satisfaction of customers. Further extensions of continuous approximation methods to estimate other performance measures, such as service level, are quite difficult. In summary, simulation is a flexible method to evaluate performance measures through a statistical approach. The disadvantage of the simulation method is that it is time consuming when compared to the application of a mathematical function. The mathematical function is a fast and systematic method. However, approaches proposed in past research are not sufficiently accurate, and they present difficulties in addressing various issues in parcel transportations services and estimating various performance measures. A systematic model is required to estimate various performance measures efficiently. The proposed stochastic model in this thesis is built to fill these gaps.

3. Analysis of Parcel Delivery Services using a Markov Model

3.1. Overview

The objective of this chapter is to evaluate business performance in terms of transportation costs and service levels. The estimation model is formulized based on the following assumptions.

- The customers' demand appears as a Poisson process with constant rate λ, and customers' locations are independently and uniformly distributed in the service region. New customers are queued in a waiting list for delivery services. (Bertsimas and Ryzin, 1991)
- There is a single vehicle with infinite capacity which operates parcel delivery services in a certain region. Once free, the vehicle plans a delivery trip which originates from the warehouse, visits customers who are in the waiting list, and returns to the warehouse.
- The loading and unloading time at the warehouse and customer locations can be ignored, since it is relatively short compared with the vehicle travel time. (Haghani and Jung, 2005)
- To simplify the problem, the vehicle serves customers according to the first-come-first-serve (FCFS) rule. (Du et al. 2005) This assumption can be relaxed in a later chapter.
- The vehicle is travelling at a constant speed of one unit distance per unit time. (Larsen 2000)

In this chapter, the author investigates the process of dynamic delivery services in a Markov model. Before constructing the Markov model, the vehicle travel time should be estimated by the continuous approximation model mentioned in chapter 2. In section 3.2, it is assumed that the vehicle travel time is approximated by a random variable with a certain distribution. Based on this assumption, delivery services are modelled as a Markov chain. In section 3.3, the transportation cost is estimated based on the steady state process of the Markov model, and the customer waiting time for the service and the service level are estimated based on a transient state process, in section 3.4. In section 3.5, a strategy to manage the vehicle departure time is investigated, and the Markov model is modified accordingly. In section 3.6, numerical results for validating the proposed model are discussed.

3.2. Approximation of Vehicle Travel Time

A single delivery trip consists of several paths. If the service sequence in the trip is ($n_1, n_2, \cdots, n_j$), the vehicle starts from the warehouse, visits n_1 followed by n_2, and so on. After the vehicle leaves n_j, it heads back to the warehouse. Let d_0 denote the period of time that the vehicle spends travelling between the warehouse and the customer n_1, d_1 be the period of time the vehicle travels between the customer n_1 and n_2, and so on. There are $j+1$ paths in this trip, and the total travel time of this trip is $l = \sum_{i=0}^{j} d_i$. In the DVRP, d_i can be considered a random variable, and the mean and variance of l can be estimated by the continuous approximation model. The expected value of d_i ($i \geq 1$) is denoted as τ_1, and its variance is σ_1^2. The mean and variance of d_0 is τ_0 and σ_0^2, respectively. Christofides and Eilon (1969) estimated that the expected time of a trip which spans N customers is proportional to the square root of the area of the service region, provided that the N customers are independently and uniformly scattered in the region. If two independently and uniformly distributed points X_1 and X_2 in a region, and the centre point X^* of the region are provided, Equations (3.1), (3.2), (3.3) and (3.4) can be formulated.

$$\tau_1 = \mathrm{E}\left[\left\|X_1 - X_2\right\|\right] = c_1\sqrt{A} \tag{3.1}$$

$$\tau_1^2 + \sigma_1^2 = \mathrm{E}\left[\left\|X_1 - X_2\right\|^2\right] = c_2 A \tag{3.2}$$

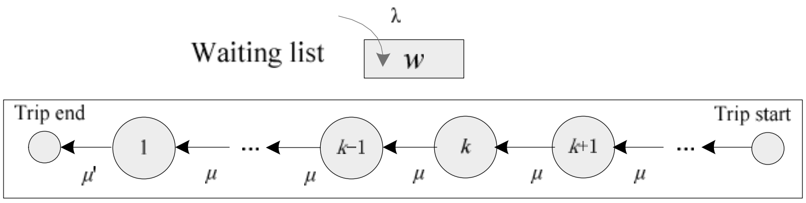
$$\tau_0 = \mathrm{E}\left[\left\|X_1 - X^*\right\|\right] = c_3\sqrt{A} \tag{3.3}$$

$$\tau_0^2 + \sigma_0^2 = \mathrm{E}\left[\left\|X_1 - X^*\right\|^2\right] = c_4 A \tag{3.4}$$

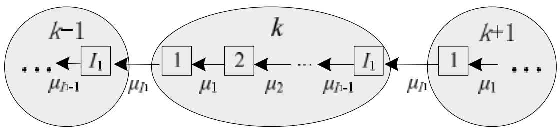
Larson and Odoni (1981) estimated that $c_1 \approx 0.52$, $c_2 \approx 1/3$, $c_3 \approx 0.383$ and $c_4 \approx 1/6$ if the service region is square. They also estimated these parameters in regions of varying shapes. Based on this estimate, the vehicle travelling between customers can be considered a stochastic process with general distribution. In dynamic delivery services, Equations (3.3) and (3.4) can be used to estimate the mean and variance of the travel time from the warehouse to the first customer, and Equations (3.1) and (3.2) can be used to estimate the travel time between any two successive customers.

Dynamic parcel delivery services can be modelled as an M/G/1 queuing system. In this system, the customer inter-arrival time is exponentially distributed, the service time follows a general distribution, and there is only one server (the vehicle) in the system. In a Markov chain, a complete specification of a system is collectively named as the state. The system state is generally defined as the number of customers N currently in the system, which means that N customers' orders have not yet been fulfilled. N is a random variable of time, it is thus denoted as N_t. Once a new customer's demand appears, the system state increases by one ($N_{t+\delta t} = N_t + 1$). Once a customer demand is fulfilled, the system state decreases by one ($N_{t+\delta t} = N_t - 1$). At any given point in time, the system must be completely described as being in one of the discrete states.

In dynamic delivery services, customers are queued in a waiting list. Once the vehicle is idle at the warehouse, a schedule for the customers is determined accordingly and the vehicle initiates a trip in order to fulfil the demands. Therefore, customers in the system have two different statuses. One status is when the customer is ready to receive the delivery according to the schedule. The other status is when the customer is queued in the waiting list with no set schedule. Therefore, the system state is defined as (*w*, *k*), where *w* is the number of customers queued in the waiting list, and *k* is the number of customers left on the current scheduled trip. $N = w + k$. The Markov chain of the system is shown in Fig.3.1(a).



a. Queuing process for dynamic parcel delivery services



b. Approximate the vehicle travelling process by hypo-exponentially distributed process

Fig. 3.1. Markov Model for parcel delivery services

The Markov property states that the probability that the system will be in a particular state at the next moment of time depends only on the current state. It requires that the transition time from one state to the other follows an exponential distribution. However, the vehicle travel time between customers follows a general distribution, which does not satisfy the Markov property requirement. In order to model dynamic parcel delivery services as a continuous time Markov Chain, the generally distributed vehicle travel time is approximated by a combination of a series of exponential distributed travel time. The total time spent on several sequential Poisson events with different rates is measured by a hypo-exponential distribution, which is used to approximate the general distribution of vehicle travel time. In order to describe the series of Poisson events, a system state (w, k) is further divided into several sub-states (w, k, I), where I represents the index of sub-states. In particular, state (0,0,0) indicates that the vehicle has become idle due to no pending customer demand to be fulfilled. If the length of the queue is limited at l customers and the number of sub-states is I_0, the number of states in the system is $I_0*l*(l+1)$. The approximation of the hypo-exponentially distributed process is illustrated in Fig.3.1(b).

The hypo-exponential distribution consists of I_0 sub-state phases for the paths between the warehouse and the first customer or the path between the last customer and the warehouse, and I_1 sub-state phases for the paths in between. Each sub-state phase is an exponential phase with a transition rate μ_i. Therefore, the mean of the hypo-exponential distribution for I_0 sub-state phases is $\sum_{i=1}^{I_0}\frac{1}{\mu_i}$, and the variance is $\sum_{i=1}^{I_0}\frac{1}{\mu_i^2}$. In order to approximate the vehicle travel time with mean τ and variance σ^2, this hypo-exponential distribution must satisfy the two equations in Equation (3.5).

$$\begin{cases} \tau = \sum_{i=1}^{I_0} \frac{1}{\mu_i} \\ \sigma^2 = \sum_{i=1}^{I_0} \frac{1}{\mu_i^2} \end{cases} \tag{3.5}$$

Proposition 3.1: If there are positive real solutions for Equation (3.5), the following condition (3.6) must hold.

$$\begin{cases} \sigma < \tau \\ \sigma \geq \tau / \sqrt{I_0} \end{cases} \tag{3.6}$$

Proof: This proposition assumes that Equation (3.5) has positive real solutions. The following mathematical induction is used to prove that condition (3.6) must hold.

(I) Let $\sigma_I = \sigma$, $\tau_I = \tau$ and $r_i = 1/\mu_i$

$\sigma_I > 0$, $\tau_I > 0$

(II) If $I_0 = 2$, the solution to Equation (3.5) is shown in Equation (3.7).

$$r_{1,2} = \frac{\tau_2}{2} \pm \sqrt{\frac{1}{2}\sigma_2^2 - \frac{1}{4}\tau_2^2} \tag{3.7}$$

If r_1 and r_2 is the positive real solution, the following must hold.

$$\begin{cases} \sigma_2 < \tau_2 \\ \sigma_2 \geq \tau_2 / \sqrt{2} \end{cases}$$

(III) It is assumed that the proposition holds when $I_0 = k - 1$ (where *k* is an integer and $k \geq 3$), which is described as follows.

$$\begin{cases} \tau_{k-1} = \sum_{i=1}^{k-1} r_i \\ \sigma_{k-1}^2 = \sum_{i=1}^{k-1} r_i^2 \end{cases}$$

The equations have positive real solutions only in the case when

$$\begin{cases} \sigma_{k-1} < \tau_{k-1} \\ \sigma_{k-1} \geq \tau_{k-1} \big/ \sqrt{k-1} \end{cases}$$

Let $\tau_k = \tau_{k-1} + r_k$ and $\sigma_k^2 = \sigma_{k-1}^2 + r_k^2$.

$$\begin{cases} \tau_k - r_k = \sum_{i=1}^{k-1} r_i \\ \sigma_k^2 - r_k^2 = \sum_{i=1}^{k-1} r_i^2 \end{cases} \tag{3.8}$$

According to the assumption when $I_0 = k-1$, the following must hold.

$$\begin{cases} \sqrt{\sigma_k^2 - r_k^2} < \tau_k - r_k \\ \sqrt{\sigma_k^2 - r_k^2} \geq (\tau_k - r_k) \big/ \sqrt{k-1} \end{cases}$$

Rearranging the above inequations while maintaining $\tau_k - r_k = \tau_{k-1} > 0$,

$$\begin{cases} \frac{1}{2}\sigma_k^2 - \frac{1}{4}\tau_k^2 < \frac{1}{4}\tau_k^2 + (r_k - \tau_k) r_k < \frac{1}{4}\tau_k^2 \\ \frac{k-1}{k}\sigma_k^2 - \frac{k-1}{k^2}\tau_k^2 \geq \left(\frac{1}{k}\tau_k - r_k\right)^2 \geq 0 \end{cases} \tag{3.9}$$

Therefore, the two following conditions hold.

$$\begin{cases} \sigma_k < \tau_k \\ \sigma_k \geq \tau_k \big/ \sqrt{k} \end{cases}$$

The proposition is proved in the situation of $I_0 = k$, provided that the proposition holds when $I_0 = k-1$. Since it has been proved that the proposition holds when $I_0 = 2$, the proposition is confirmed in all conditions when $I_0 \geq 2$.

Since τ and σ are proportional to $\sqrt{A}$, and all μ_i must satisfy Equation (3.5), μ_i is proportional to $1/\sqrt{A}$. Therefore, it is not necessary to seek feasible solutions to Equation (3.5) again, when delivery services have the same region shape but different sizes. The solution to Equation (3.5) is as follows.

According to condition (3.9), the range of r_k can be calculated.

Case I. $\frac{\tau_k}{\sqrt{k}} > \sigma_k \geq \frac{\tau_k}{\sqrt{k-1}}$

$$\frac{\tau_k - \sqrt{k(k+1)\sigma_k^2 - k\tau_k^2}}{k+1} \leq r_k \leq \frac{\tau_k + \sqrt{k(k+1)\sigma_k^2 - k\tau_k^2}}{k+1}$$

Case II. $\frac{\tau_k}{\sqrt{2}} > \sigma_k \geq \frac{\tau_k}{\sqrt{k}}$

$$0 < r_k \leq \frac{\tau_k + \sqrt{k(k+1)\sigma_k^2 - k\tau_k^2}}{k+1}$$

Case III. $\tau_k > \sigma_k \geq \frac{\tau_k}{\sqrt{2}}$

$$0 < r_k \leq \frac{\tau_k - \sqrt{2\sigma_k^2 - \tau_k^2}}{2} \quad \text{or}$$

$$\frac{\tau_k + \sqrt{2\sigma_k^2 - \tau_k^2}}{2} \leq r_k \leq \frac{\tau_k + \sqrt{k(k+1)\sigma_k^2 - k\tau_k^2}}{k+1}$$

In order to find one group of feasible solutions for Equation (3.5), the following algorithm is proposed.

Step I, check to see if $\tau \geq \sigma$. If so, there are feasible solutions. Since the fewer states has the system, the faster result is obtained from the calculation, the value of I_0 is determined by the minimum integer of $I_0 \geq \tau^2 / \sigma^2$.

Step II, let $\sigma_k = \sigma$, $\tau_k = \tau$ and $r_i = 1/\mu_i$. Decide the range of r_k based on Case I to Case III. Assign a random value to r_k within the range.

Step III, let $\tau_{k-1} = \tau_k - r_k$, $\sigma_{k-1}^2 = \sigma_k^2 - r_k^2$ and $k = k-1$. Repeat Step II and III until $k = 2$.

Step IV, use Equation (3.7) to find the value of r_1 and r_2.

Step V, obtain μ_i from the inverse value of r_i.

For example, suppose there is a square region with area *A*, and there is a warehouse in the middle of the region. The vehicle plans the travelling route based on the FCFS rule. The travel distance between customers and the warehouse can be estimated by the results proposed by Larson and Odoni (1981). Let τ_1 denote the average travel time between two successive customers' locations, σ_1^2 be the variance of these travel time, τ_0 denote the average travel time between a customer location and the warehouse and σ_0^2 be the variance. Therefore, $\tau_1 = 0.52\sqrt{A}$, $\sigma_1^2 = A/3 - \tau_1^2 = 0.0629A$, $\tau_0 = 0.383\sqrt{A}$ and $\sigma_0^2 = 0.02A$. According to Equation (3.5), the following equations need to be solved.

$$\begin{cases} \sum_{i=1}^{I_1} \dfrac{1}{\mu_i} = 0.52\sqrt{A} \\ \sum_{i=1}^{I_1} \dfrac{1}{\mu_i^2} = 0.0629A \\ \sum_{i=1}^{I_0} \dfrac{1}{\mu_i'} = 0.383\sqrt{A} \\ \sum_{i=1}^{I_0} \dfrac{1}{\mu_i'^2} = 0.02A \end{cases}$$

where I_0 and I_1 are restricted by condition (3.6).

$I_1 \ge \tau_1^2/\sigma_1^2 = 4.3$ and $I_0 \ge \tau_0^2/\sigma_0^2 = 7.3$

Following the algorithms provided above, one group of solutions for μ_i and μ_i' is obtained.

$\mu_i = \{32.6, 10, 9.1, 8.3, 6.3\} / \sqrt{A}$

$\mu_i' = \{73.6, 23, 21, 20.8, 19.2, 17.9, 16.8, 15.9\} / \sqrt{A}$

In this case, the number of states in the system is (8+14*5)*(15+1) = 1248, if the length of the queue is set at 15 customers.

3.3. Transportation Cost Estimation

This section presents methods for estimating transportation costs, which are assumed to be proportional to the average vehicle travel distance within a specific period. Since the vehicle speed is constant and the status of the vehicle is either travelling or idle, the average travel distance is proportional to the percentage of time that the vehicle spends travelling, which is defined as the vehicle utilization. The higher the vehicle utilization, the higher the transportation costs. Vehicle utilization can be calculated based on a steady state process of the Markov Chain, which is introduced as follows.

In general, the behavior of a Markov Chain is investigated when the system is in steady state. In this situation, the probability that the system is in a certain state is independent of time and the starting state. This probability is named steady state probability, $\pi = \{\pi_v\}$, where π_v is the steady state probability of state *v*. The steady state probability is obtained by resolving balance equations, in which the average total flow into a certain state is equal to the average total flow out of the state.

1) The vehicle is travelling between customers. (State (*w*,*k*,*I*))

   Fig. 3.2 illustrates the transition diagram when the vehicle is travelling between customers' locations. The system state currently is (*w*,*k*,*I*), and arrows show the probability flows into and out of this state.

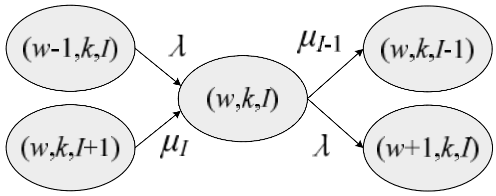


Fig. 3.2. Transition diagram of the vehicle travelling between customers

Based on the transition diagram in Fig. 3.2, the balance equation is formulated as Equation (3.10)

$$\mu_I \pi_{(w,k,I+1)} + \lambda \pi_{(w-1,k,I)} = (\mu_{I-1} + \lambda)\pi_{(w,k,I)} \tag{3.10}$$

The left side of Equation (3.10) represents the total probability of transitioning into state (w,k,I) from other states. The right side of the equation indicates the total probability of transitioning from state (w,k,I) to other states. The transition occurs when the vehicle continues travelling or a new customer appears.

2) Vehicle reaching and leaving a customer's location. (State (w,k,I_1))

When the vehicle reaches a customer's location, it starts heading to the next customer immediately, due to the assumption of zero unloading time at customer's location. Fig. 3.3 illustrates the transition diagram in this situation.

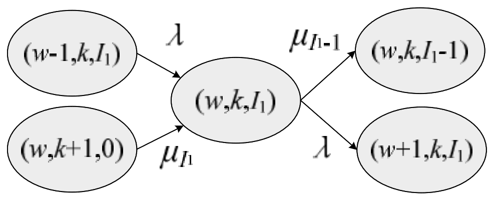


Fig. 3.3. Transition diagram of the vehicle reaching and leaving customers' locations

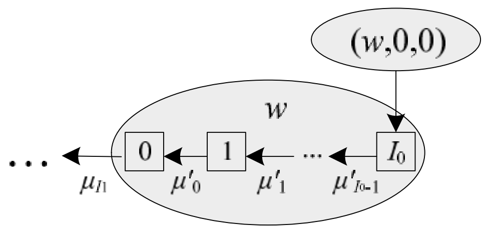
Based on the transition diagram, the balance equation is formulated as follows.

$$\mu_{I_1} \pi_{(w,k+1,0)} + \lambda \pi_{(w-1,k,I_1)} = (\mu_{I_1-1} + \lambda)\pi_{(w,k,I_1)} \tag{3.11}$$

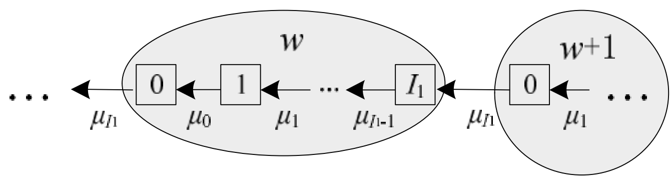
3) The vehicle returns to the warehouse and starts the next trip (Redundant state (w,0,0))

Once the vehicle finishes the current trip, the system state is expected to reach (w,0,0) from (w,0,1). The vehicle should begin travelling to the pending customer immediately, since the vehicle will spend no time loading and unloading products at the warehouse. Therefore, the vehicle starts the next trip once it returns to the warehouse in this case, and the process transitions from state (w,0,1) directly to state (0,w,0).

When the vehicle starts a new trip, the vehicle should move from the warehouse to the first customer, which is illustrated in Fig.3.4.(a). In this process, the flow rate between two system states is μ_i'. Fig.3.4.(b) illustrates the process that the vehicle travels from the w+1 customer to the w customer with a flow rate of μ_i between system states. However, these two processes share the same series of transitions from state (0,w,I) to state (0,w,0). It is difficult to construct these two processes with different transition rates in the same Markov chain.



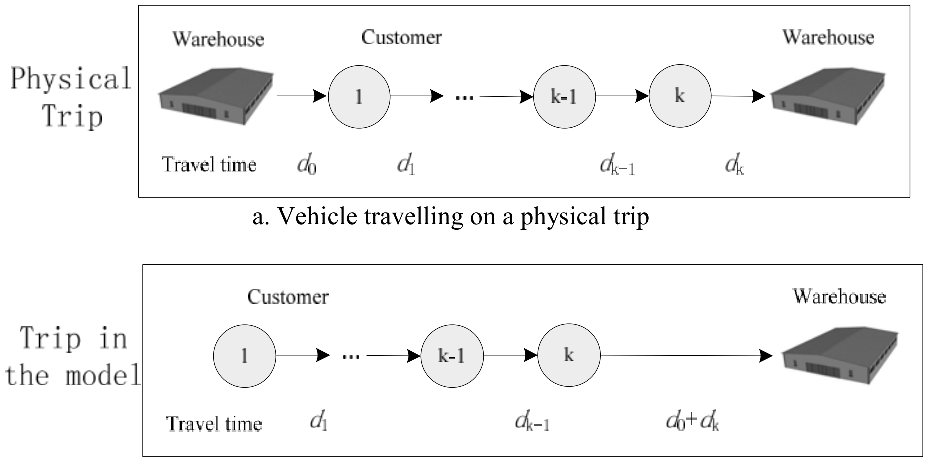
a. Transportation from the warehouse to the first customer in a trip.



b. Transportation from the w+1 customer to the w customer

Fig. 3.4. The difference between the process from the warehouse to customer and the process from customer to customer.

In order to overcome this difficulty, the process of the vehicle travelling from the warehouse to the first customer and the process of the vehicle travelling from the last customer to the warehouse are combined, which are illustrated in Fig. 3.5. The transition rate from state (w,0,I) to state (w,0,I-1) becomes $\mu'_{I-1}/2$, due to this combined process. This combined process is at the end of each trip, and the vehicle starts the trip by travelling from the first customer to the second customer. Although this mechanism rearranges the physical vehicle travelling sequence, it will not affect the final results, since only vehicle utilization is examined in this section.



a. Vehicle travelling on a physical trip

b. Vehicle travelling process in the Markov model

Fig. 3.5. Modeling the vehicle travelling process on a trip

Based on the transition diagram in Fig. 3.6, the balance equation is formulated as Equation (3.12).

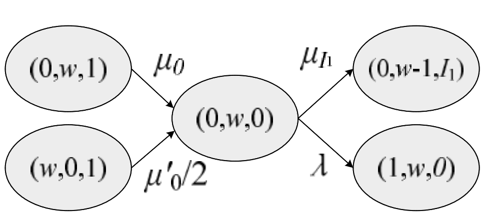


Fig. 3.6. Transition diagram in the situation that the vehicle returns to the warehouse and starts the next trip

$$\mu_0' \pi_{(w,0,1)}/2 + \mu_0 \pi_{(0,w,1)} = (\mu_{I_1} + \lambda)\pi_{(0,w,0)} \tag{3.12}$$

The left side of Equation (3.12) represents the transition of the vehicle travelling and returning to the warehouse. The right side of Equation (3.12) indicates the transition out of the state (0,w,0), which happens when the vehicle starts the next trip or a new customer demand appears.

4) The vehicle travels from the last customer to the warehouse on the current trip. (State (w,0,I))

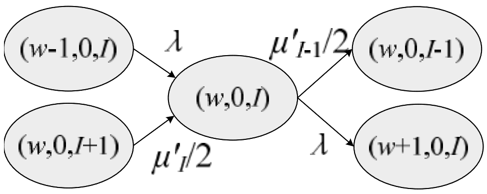


Fig. 3.7. Transition diagram in the situation that the vehicle travels towards the warehouse

As previously mentioned, the transition rate of the vehicle travelling in this case becomes $\mu_{I-1}'/2$, due to the combined process. Based on the transition diagram above, the balance equation is as follows.

$$\mu_I' \pi_{(w,0,I+1)}/2 + \lambda \pi_{(w-1,0,I)} = (\mu_{I-1}'/2 + \lambda)\pi_{(w,0,I)} \tag{3.13}$$

5) The vehicle is idle on the warehouse. (State (0,0,0))

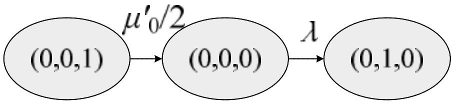


Fig. 3.8. Transition diagram of vehicle idle at the warehouse

When there are no customers waiting for services, the vehicle is idle at the warehouse. Fig. 3.8 illustrates the situation. The balance equation for this situation is as follows.

$$\mu_0' \pi_{(0,0,1)}/2 = \lambda \pi_{(0,0,0)} \tag{3.14}$$

The balance equations can be summarized as the following stationary equation.

$$\pi Q = 0 \tag{3.15}$$

$Q = \{q_{u,v}\}$ is defined as the intensity matrix, and $q_{u,v}$ specifies the transition rate from state u to v. Each element in the intensity matrix is elaborated as follows.

At any moment, the state may transition from (w,k,I) to (w,k,I-1) with a flow rate μ_{I-1}, where the vehicle continues travelling between customers.

$$q_{(w,k,I),(w,k,I-1)} = \mu_{I-1}, \quad w = 0,1,\cdots; k = 1,2,\cdots; I = 1,2,\cdots,I_1$$

The flow rate is $\mu'_{I-1}/2$, when the vehicle travels between a customer and the warehouse.

$$q_{(w,0,I),(w,0,I-1)} = \mu'_{I-1}/2, \quad w = 0,1,\cdots; I = 2,3,\cdots,I_0$$

Due to the redundant state (w,0,0), the state may transition directly from (w,0,1) to (0,w,0) with the flow rate $\mu'_0/2$, when the vehicle returns to the warehouse and starts the next trip.

$$q_{(w,0,1),(0,w,0)} = \mu'_0/2, w = 1,2,\cdots$$

The system stays in the (0,0,0) state, if there are no customers waiting for the service.

$$q_{(0,0,1),(0,0,0)} = \mu'_0/2$$

The system transitions from state (w,k,0) to (w,k-1,I_1) with flow rate μ_{I_1}, when the vehicle finishes the service for a customer and sets out for the next destination.

$$q_{(w,k,0),(w,k-1,I_1)} = \mu_{I_1}, \quad w = 0,1,\cdots; k = 2,3,\cdots$$

The system transitions from state (w,1,0) to (w,0,I_0) with flow rate μ'_{I_0}, when the vehicle finishes the service for the last customer on current trip and starts heading back to the warehouse.

$q_{(w,1,0),(w,0,I_0)} = \mu'_{I_0}/2$, $w = 0,1,\cdots$

When a new customer appears, the number of customers in the system increases by one.

$q_{(w,k,I),(w+1,k,I)} = \lambda$, $w = 0,1,\cdots; k = 0,1,\cdots; I = 0,1,\cdots$, (except $q_{(0,0,0),(1,0,0)}$)

If a new customer appears when the vehicle is sitting idle at the warehouse, the vehicle will immediately set out to the location of this new customer. The state transitions from (0,0,0) to (0,1,0) and the flow rate λ can be determined.

$q_{(0,0,0),(0,1,0)} = \lambda$

Based on the definition of the intensity matrix, the flow rates out of a certain state should add up to 0. Therefore, the diagonal elements of the intensity matrix can be calculated as follows.

$q_{(w,k,I),(w,k,I)} = -\lambda - \mu_{I-1}$, $w = 0,1,\cdots; k = 1,2,\cdots; I = 1,2,\cdots$

$q_{(w,0,I),(w,0,I)} = -\lambda - \mu'_{I-1}/2$, $w = 0,1,\cdots; I = 1,2,\cdots$

$q_{(w,k,0),(w,k,0)} = -\lambda - \mu_{I_1}$, $w = 0,1,\cdots; k = 2,3,\cdots$

$q_{(w,1,0),(w,1,0)} = -\lambda - \mu'_{I_0}/2$, $w = 0,1,\cdots$

$q_{(0,0,0),(0,0,0)} = -\lambda$

The steady state probability π is obtained by the stationary equation together with the boundary condition $\sum \pi = 1$.

$$\begin{cases} 0 = \pi Q \\ \sum \pi = 1 \end{cases} \tag{3.16}$$

The vehicle utilization U is obtained from the following equation.

$$U = \sum_{u \neq (0,0,0)} \pi_u = 1 - \pi_{(0,0,0)} \tag{3.17}$$

$\sum_{u \neq (0,0,0)} \pi_u$ is the steady state probability of the vehicle travelling and $\pi_{(0,0,0)}$ is the steady state probability of the vehicle sitting idle at the warehouse.

The transportation cost per unit time C_T is proportional to the vehicle utilization.

$$C_T = \varphi_1 U = \varphi_1 (1 - \pi_{(0,0,0)}) \tag{3.18}$$

where φ_1 is the petrol cost per unit time when vehicle is travelling.

3.4. Service Level Estimation

The customer waiting process discussed in this section is to estimate the waiting time for a customer. This specific customer represents a customer living anywhere in the region and making an order at any time. It is assumed that this process starts from a setting time 0, when the system is stable and the specific customer appears. It ends when the customer's demand has been satisfied and the customer is no longer in the system. Therefore, the probability of each state continues changing with time, and the system will never become stable. In other words, this process is a transient state process. The purpose is to track the entire process of a customer from his arrival until the completion of the service for him/her, in order to calculate the span of time the customer stays in the system.

There are two vehicle routing trips in this process. The first trip is the current trip, when the specific customer appears. After the first trip, the vehicle plans the next trip for the specific customer and the remaining customers in the waiting list. The variance of state probabilities is investigated along the time axis. Finally, the distribution of customer waiting time in the system is obtained. With this distribution, service levels can be defined based on different criterions.

In this process, the Markov model needs to track the position of the specific customer in the queue once he or she appears. The position of the specific customer is an important point of reference, which need to be added into the definition of the state. The system state is now defined as (w,k,I,b), where the first three symbols have the same meaning as the previous steady state process, and b indicates the position of the specific customer in the queue. However, in this four-dimension state, the state space will exceed ten thousand if each dimension has more than ten different values. Solving a Markov model with a large state space will cost considerable computational efforts. In order to efficiently estimate customer waiting time, reducing the state space is necessary.

In this process, the service sequence for the next vehicle trip is fixed once the specific customer appears, due to the FCFS queue discipline. The demand after this specific customer will not affect the final results. It is not necessary to track the number of customers queued in the waiting list, after the service sequence of the specific customer on the next trip is determined. The definition of the state in this process can be reduced to three dimensions (k,I,b). b does not change until the vehicle finished the current trip and starts the next trip. After that, b is set to 0, in order to indicate the process for the second trip. In particular, The state (0,0,0) indicates that the demand of the specific customer has been fulfilled. Let $\pi'_t = \{\pi'_{u,t}\}$ be the transient state probability at time t, where $\pi'_{u,t}$ is the transient state probability of state u at time t.

Initially, at time 0, the specific customer appears and joins the end of the queue in the waiting list. The initial state probability $\pi'_{(k,I,b),0}$ is determined by the steady state probability π obtained from previous process, due to the assumption that the system is stable at time 0. It can be proved that the system is in a certain state with the same probability as the previous steady state probability, when the specific customer appears, due to the assumption that customers appear in a Poisson manner (see the book edited by Gross and Harris, 1998, page 221-222).

$$\pi'_{(k,I,b),0} = \pi_{(b-1,k,I)}, \quad k = 0,1,\cdots; I = 0,1,\cdots; b = 1,2,\cdots, \text{(except } \pi'_{(0,0,1),0}\text{)} \tag{3.19}$$

If the vehicle is initially idle, the specific customer will be served immediately. The initial probabilities in other states are set to zero.

$$\pi'_{(1,0,0),0} = \pi_{(0,0,0)} \tag{3.20}$$

In the transient state process of the Markov Chain, the state probability changes due to the difference between rates flowing into the state and rates flowing out off the state.

1) The vehicle travels between customers. (State (k,I,b))

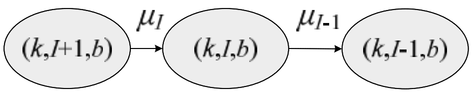


Fig. 3.9. Transition diagram of vehicle travelling between customers

Based on the transition diagram in Fig. 3.9, the transient state equation is formulated as Equation (3.21)

$$\frac{d\pi'_{(k,I,b),t}}{dt} = \mu_I \pi'_{(k,I+1,b),t} - \mu_{I-1} \pi'_{(k,I,b),t} \tag{3.21}$$

The first element on the right side of Equation (3.21) represents the transition rates from other states into state (k,I,b). The second element on the right side of the equation indicates the transition rates out off state (k,I,b).

2) The vehicle reaches and leaves a customer's location. (State (k,I_1,b))

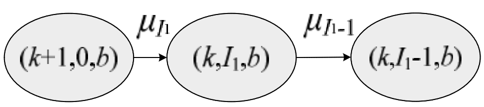


Fig. 3.10. Transition diagram of a vehicle reaching and leaving customers' locations

$$\frac{d\pi'_{(k,I_1,b),t}}{dt} = \mu_{I_1}\pi'_{(k+1,0,b),t} - \mu_{I_1-1}\pi'_{(k,I_1,b),t} \tag{3.22}$$

3) The vehicle returns to the warehouse and starts the next trip. (State (b,0,0))

As previously mentioned, the state (0,0,b) is redundant. The vehicle starts the second trip once it returns to the warehouse and the process transitions from state (0,1,b) directly to state (b,0,0).

Since this process ends when the vehicle reaches the specific customer's location, the vehicle's return trip can be ignored. Therefore, the second trip is slightly different from the one shown in Fig. 3.5. The transition rate from state (0,I,0) to state (0,I-1,0) is μ'_{I-1}. The second trip can be reconstructed with the initial path from the first customer to the second and it ends with the path from the warehouse to the first customer. Although this mechanism rearranges the physical vehicle travelling sequence, it will not affect the final results, since analyzing the customer waiting time is the only objective in this section.

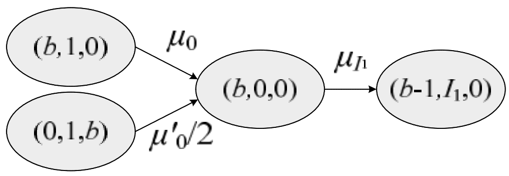


Fig. 3.11. Transition diagram in the situation that the vehicle returns to the warehouse and starts the next trip

Based on the transition diagram in Fig. 3.11, the transient state equation is formulated as Equation (3.23).

$$\frac{d\pi'_{(b,0,0),t}}{dt} = \frac{\mu'_0}{2}\pi'_{(0,1,b),t} + \mu_0\pi'_{(b,1,0),t} - \mu_{I_1}\pi'_{(b,0,0),t} \tag{3.23}$$

4) The vehicle travels to the warehouse. (State (0,I,b))

As previously mentioned, the first trip in the customer waiting process is a circular trip which starts and ends at the warehouse. In the Markov model, the last leg of the first trip includes a combination of two journeys between the customer and the warehouse (Fig. 3.12 (a)). This process takes an incomplete second trip into consideration, which only includes one journey from the warehouse to the first customer. Therefore, the transition rate in the last step of the second trip is μ'_{I-1}, which is illustrated in Fig. 3.12 (b).

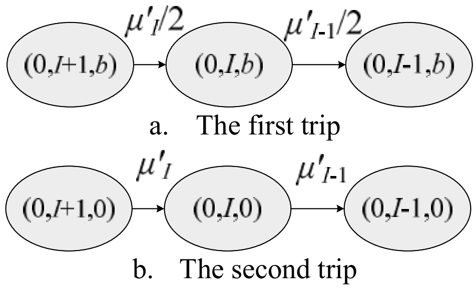


Fig. 3.12. Transition diagram in the situation that the vehicle is travelling towards the warehouse

Based on the transition diagram of Fig. 3.12, the transient state equations are Equation (3.24) for the first trip and Equation (3.25) for the second trip.

$$\frac{d\pi'_{(0,I,b),t}}{dt} = \frac{\mu'_I}{2}\pi'_{(0,I+1,b),t} - \frac{\mu'_{I-1}}{2}\pi'_{(0,I,b),t} \tag{3.24}$$

$$\frac{d\pi'_{(0,I,0),t}}{dt} = \mu'_I\pi'_{(0,I+1,0),t} - \mu'_{I-1}\pi'_{(0,I,0),t} \tag{3.25}$$

5) The vehicle finishes the service for the specific customer. (State (0,0,0))

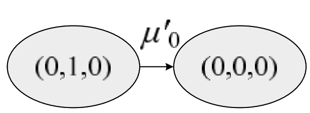


Fig. 3.13. Transition diagram of service finished

The transient state equation for this situation is as follows.

$$\frac{d\pi'_{(0,0,0),t}}{dt} = \mu'_0\pi'_{(0,1,0),t} \tag{3.26}$$

The above differential equations can be summarized as the following forms.

$$\frac{d\pi_t'}{dt} = \pi_t' Q'$$

Resolving this differential equation, the following solution is obtained.

$$\pi_t' = \pi_0' e^{tQ'} \tag{3.27}$$

where $Q' = \{q'_{u,v}\}$ is the intensity matrix for this process, and $q'_{u,v}$ is the transition rate from state u to v, which is illustrated as follows.

The transition rate from (k,I,b) to (k,I-1,b) is μ_{I-1}, when the vehicle is travelling between customers.

$$q'_{(k,I,b),(k,I-1,b)} = \mu_{I-1}, \quad k = 1,2,\cdots; I = 1,2,\cdots; b = 0,1,\cdots$$

On the first trip the transition rate is $\mu'_{I-1}/2$, when the vehicle travels between the warehouse and customer.

$$q'_{(0,I,b),(0,I-1,b)} = \mu'_{I-1}/2, \quad I = 2,3,\cdots; b = 1,2,\cdots$$

Due to the redundant state (0,0,b), the state may transition directly from (0,1,b) to (b,0,0) with the flow rate $\mu_0'/2$, when the vehicle returns to the warehouse and starts the second trip.

$$q'_{(0,1,b),(b,0,0)} = \mu_0'/2, \quad b = 1,2,\cdots$$

On the second trip the transition rate is μ'_{I-1}, when the vehicle is travelling between the warehouse and customer.

$$q'_{(0,I,0),(0,I-1,0)} = \mu'_{I-1}, \quad I = 1,2,\cdots$$

The system transitions from state (k,0,b) to (k-1,I_1,b) with transition rate μ_{I_1}, when the vehicle finishes the service for a customer and sets out for the next destination.

$$q'_{(k,0,b),(k-1,I_1,b)} = \mu_{I_1}, \quad k = 2,3,\cdots; b = 0,1,\cdots$$

The system transitions from state (1,0,b) to (0,I_0,b) with transition rate $\mu'_{I_0}/2$, when the vehicle finishes the service for the last customer on first trip and starts heading to the warehouse.

$$q'_{(1,0,b),(0,I_0,b)} = \mu'_{I_0}/2, \quad b = 1,2,\cdots$$

The transition rate is μ'_{I_0}, when the vehicle starts the last step of the second trip.

$$q'_{(1,0,0),(0,I_0,0)} = \mu'_{I_0}$$

When the vehicle finishes the delivery for the specific customer, the process terminates at state (0,0,0).

$$q'_{(0,1,0),(0,0,0)} = \mu'_0$$

The diagonal elements of the intensity matrix are calculated as follows.

$$q'_{(k,I,b),(k,I,b)} = -\mu_{I-1}, \quad k = 1,2,\cdots; I = 1,2,\cdots; b = 0,1,\cdots$$

$$q'_{(k,0,b),(k,0,b)} = -\mu_{I_1}, \quad k = 2,3,\cdots; b = 0,1,\cdots$$

$$q'_{(1,0,b),(1,0,b)} = -\mu'_{I_0}/2, \quad b = 1,2,\cdots$$

$$q'_{(1,0,0),(1,0,0)} = -\mu'_{I_0}$$

$$q'_{(0,I,b),(0,I,b)} = -\mu'_{I-1}/2, \quad I = 1,2,\cdots; b = 1,2,\cdots$$

$$q'_{(0,I,0),(0,I,0)} = -\mu'_{I-1}, \quad I = 1,2,\cdots$$

$$q'_{(0,0,0),(0,0,0)} = 0$$

The solution of $\pi'_{(0,0,0),t}$ can be obtained at any time t based on Equation (3.27), subject to the initial conditions (3.19) and (3.20). $\pi'_{(0,0,0),t}$ is the cumulative distribution function of the waiting time of the specific customer. Finally, setting the criterion of customer waiting time as T, the service level S_T is

$$S_T = \pi'_{(0,0,0),T} \times 100\% \tag{3.28}$$

A numerical technique is introduced to approximate the calculation for π'_t as follows. Firstly, the solution for a corresponding discrete time Markov chain is investigated, in which $M' = \{m'_{u,v}\}$ denotes the transition matrix.

$$M' = Q'/\Omega + \mathrm{I} \tag{3.29}$$

In Equation (3.29), I is the identity matrix, and Ω is a relatively large integer. One unit time is divided into Ω small time periods, so that more than one independent event may occur with rare chance within this small time period. In practice, Ω is the nearest integer greater than $\max_{u,v}\left|q'_{u,v}\right|$. Therefore, the elements in the transition matrix are as follows.

$$\begin{cases} m'_{u,v} = q'_{u,v}/\Omega, \forall u \neq v \\ m'_{u,u} = 1 + q'_{u,u}/\Omega, \forall u \end{cases}$$

Let V'_i be the i^{th} step transient state probability in the discrete time Markov chain. Based on the discrete time Markov chain property, $V'_{i+1} = V'_i M'$, and V'_i can be calculated with the initial condition $V'_0 = \pi'_0$.

The solution for π'_t can be obtained by the following equation.

$$\pi'_t = \sum_{i=0}^{\infty} V'_i \frac{(\Omega t)^i}{i!} e^{-\Omega t} = V'_0 \sum_{i=0}^{\infty} M'^{(i)} \frac{(\Omega t)^i}{i!} e^{-\Omega t}$$

The computation problem remains that there is an infinite summation of above equation. The sum can be truncated at some value, for example *r*, and the solution can be obtained with an acceptable error ε_r (Gross and Harris, 1998).

$$\pi'_t = V'_0 \sum_{i=0}^{r} M'^{(i)} \frac{(\Omega t)^i}{i!} e^{-\Omega t} + V'_0 \sum_{i=r+1}^{\infty} M'^{(i)} \frac{(\Omega t)^i}{i!} e^{-\Omega t}$$

$$\left\| V_0' \sum_{i=r+1}^{\infty} M'^{(i)} \frac{(\Omega t)^i}{i!} e^{-\Omega t} \right\|_{\infty} \le \sum_{i=r+1}^{\infty} \frac{(\Omega t)^i}{i!} e^{-\Omega t}$$

The error of π_t' can be bounded at the maximum value ε_r, by finding *r* such that

$$\sum_{i=r+1}^{\infty} \frac{(\Omega t)^i}{i!} e^{-\Omega t} \le \varepsilon_r$$

Based on the property of Chernoff bound (Kobayashi et al., 2012), inequation (3.30) can be used to find suitable *r* with desired bounded error.

$$\sum_{i=r+1}^{\infty} \frac{(\Omega t)^i}{i!} e^{-\Omega t} \le \frac{e^{-\Omega t}(e\Omega t)^r}{r^r} \le \varepsilon_r \text{, for } r > \Omega t \tag{3.30}$$

In the proposed model, $\Omega t \ge 100$, since the criterions in the calculations of service levels in the following experiments are set to be greater than or equal to 100 units of time. If *r* is set to be $1.25\Omega t$, the solution of Equation (3.31) is acceptable with error less than 0.056.

$$\sum_{i=1.25\Omega t+1}^{\infty} \frac{(\Omega t)^i}{i!} e^{-\Omega t} \le \left(\frac{e^{1.25-1}}{1.25^{1.25}} \right)^{\Omega t} \le 0.055413$$

$$\pi_t' = \pi_0' \sum_{i=0}^{r} (Q'/\Omega + \mathrm{I})^{(i)} \frac{(\Omega t)^i}{i!} e^{-\Omega t} \tag{3.31}$$

3.5. Issue on Vehicle Departure Strategy

In the previous discussion, it is assumed that the vehicle will start a trip once there is a customer waiting for the service. However, it may not be efficient to start the trip immediately with only a small number of customers in the waiting list. The vehicle can wait at the warehouse until there are an appropriate number of customers that can be served together on one trip, in order to save travelling costs and adjust customer waiting time properly as well. A possible vehicle departure strategy may dictate that the vehicle only start a trip with at least N_D ($N_D \geq 1$) customers waiting for delivery services. Based on the calculation in sections 3.3 and 3.4, several formulas need to be modified to adapt this newly defined vehicle departure strategy.

When $w \geq N_D$, Equation (3.12) is still valid. However, when there are not enough customers waiting at the time the vehicle returns to the warehouse ($w < N_D$), the state (w,0,0) is valid, and the situation needs to be further analyzed.

When $w < N_D - 1$, the vehicle can only wait for new demands until $w = N_D$ (Fig. 3.14).

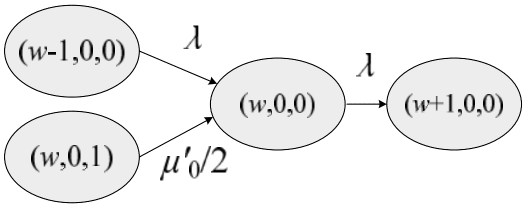


Fig. 3.14. Transition diagram of vehicle idle at the warehouse when $w < N_D - 1$

Based on the transition diagram, the balance equation is formulated as follows.

$$\mu_0' \pi_{(w,0,1)} / 2 + \lambda \pi_{(w-1,0,0)} = \lambda \pi_{(w,0,0)} \tag{3.32}$$

When $w = N_D - 1$, the vehicle immediately commences a new trip once a new customer appears (Fig. 3.15).

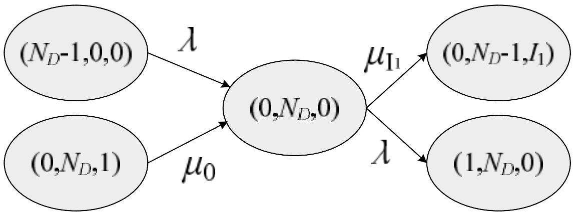


Fig. 3.15. Transition diagram of a new trip started when $w = N_D - 1$

The balance equation for this situation is as Equation (3.33).

$$\lambda \pi_{(N_D-1,0,0)} + \mu_0 \pi_{(0,N_D,1)} = (\lambda + \mu_{I_1}) \pi_{(0,N_D,0)} \tag{3.33}$$

Due to the modification of the balance equations, some elements in the intensity matrix Q are modified as follows.

$q_{(w,0,1),(w,0,0)} = \mu_0'/2$, $w = 0,1,\cdots,N_D - 1$

$q_{(w,0,1),(0,w,0)} = \mu_0'/2$, $w = N_D, N_D + 1,\cdots$

$q_{(w,k,I),(w+1,k,I)} = \lambda$, $w = 0,1,\cdots; k = 0,1,\cdots; I = 0,1,\cdots$, (except $q_{(N_D-1,0,0),(N_D,0,0)}$)

$q_{(N_D-1,0,0),(0,N_D,0)} = \lambda$

There are some elements mentioned in section 3.3 that are not included above. The vehicle utilization U is obtained by resolving the stationary equation.

$$U = 1 - \sum_{i=0}^{N_D-1} \pi_{(i,0,0)} \tag{3.34}$$

where $\sum_{i=0}^{N_D-1} \pi_{(i,0,0)}$ is the steady state probability of vehicle idle at the warehouse.

The transportation cost per unit time is formulated as Equation (3.35).

$$C_T = \varphi_1 \left(1 - \sum_{i=0}^{N_D-1} \pi_{(i,0,0)} \right) \tag{3.35}$$

In order to evaluate the service level, the transient state process has to be constructed. In this process, although the service sequence for the next vehicle trip is fixed once the specific customer appears, it is still necessary to track the number of customers in the waiting list. For example, when the specific customer appears and there are less than N_D customers in the waiting list, the vehicle has to wait until enough demand in order to start any subsequent trip. Therefore, the four dimension state (w,k,I,b) is still necessary to describe the process. The value of w will not affect the second trip, but it is a variable to track whether the system meets the criterion of vehicle departure. In order to reduce the state space, the structure of the Markov chain and the transitions between states has to be designed carefully. The initial states are grouped into four categories at time 0, when the specific customer appears.

1) There are a minimum of N_D customers (including the specific customer) in the waiting list and the vehicle is travelling between customers.

   In this situation, the demand after the specific customer will not affect the final result, since the number of customers in the waiting list is enough to trigger a second trip. It is not necessary to track the number of customers in the waiting list. Furthermore, the second trip is divided into two parts in the Markov Chain. One part can be defined as the process that the vehicle travels between a customer and the warehouse, which occurs after the first vehicle trip. The other part can be defined as the process that the vehicle travels between customers on the second trip. The second part is relocated and inserted into the first vehicle trip (an example is provided in Fig. 3.16). The reconstruction of the Markov Chain will not affect the final results because there are no differences between the first and the second trip when the vehicle is travelling between customers' locations.

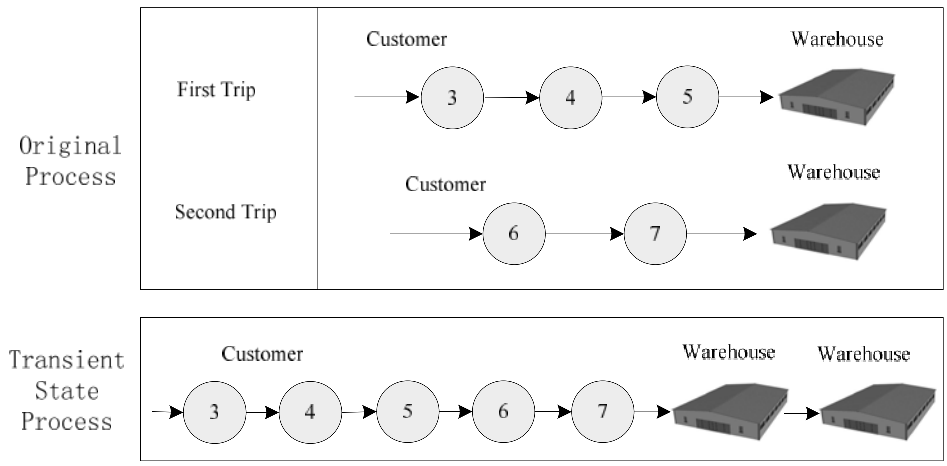


Fig. 3.16. Reconstruction of the vehicle trip in the transient state process in the situation where there are at least N_D customers in the waiting list when the specific customer appears and the vehicle is travelling between customers

In this reconstructed process, the remaining process for all situations in this category is similar, and there are no different remaining processes for different values of w if the initial state probabilities are set up properly. Therefore, the initial state for different values of w can be combined.

Initially, the same $w = N_D$ and $b = 1$ are used for all situations in this category, where b is only a notation instead of the real position of the specific customer. The initial state probabilities are formulated as follows.

$$\pi'_{(N_D,k,I,1),0} = \sum_{i \geq N_D - 1}^{k-1} \pi_{(i,k-i,I)}, \quad k = N_D, N_D + 1, \cdots ; I = 0, 1, \cdots \tag{3.36}$$

2) There are less than N_D customers (including the specific customer) in the waiting list and the vehicle is travelling between customers.

In this situation, the number of customers should be accumulated in the waiting list until it reaches N_D in order to trigger the second trip. The number of customers w in the waiting list will increase by one each time a new demand appears. After w reaches N_D, any new demand will not be taken into account, since it is at this point that the final results will no longer be affected.

b records the position of the specific customer on the second trip, and is fixed. Initially, w is equal to b. However, w will keep increasing if there are new demands added to the queue. After there are enough customers in the waiting list, no further demands will be considered in the process. The initial state probabilities are formulated as follows.

$$\pi'_{(w,k,I,b),0} = \pi_{(w-1,k,I)}, \quad w = b; k = 1, 2, \cdots ; I = 0, 1, \cdots ; b = 1, 2, \cdots, N_D - 1 \tag{3.37}$$

3) There are at least N_D customers (including the specific customer) in the waiting list and the vehicle is travelling between a customer and the warehouse.

   In this situation, the vehicle is heading to the warehouse. It reaches the warehouse and immediately starts the second trip, since there are enough customers in the waiting list. No new demands will be taken into consideration in this case. w will never change until the second trip, and b is equal to w. The initial state probabilities are as follows.

$$\pi'_{(w,0,I,w),0} = \pi_{(w-1,0,I)}, \quad w = N_D, N_D+1,\cdots; I = 0,1,\cdots \tag{3.38}$$

4) There are less than N_D customers (including the specific customer) in the waiting list and the vehicle is travelling between a customer and the warehouse.

   In this situation, b is the position of the specific customer on the second trip, and it is fixed. Initially, w is equal to b, but w will keep increasing if there are new demands. Once there are enough customers in the waiting list, no new demands will be considered in the process. The initial state probabilities are formulated in Equation (3.39).

$$\pi'_{(w,0,I,b),0} = \pi_{(w-1,0,I)}, \quad w = b; I = 0,1,\cdots; b = 1,2,\cdots, N_D - 1 \tag{3.39}$$

In the transient state process, the transition diagram and the differential equation are the same as section 3.4, when the number of customers in the waiting list is greater than or equal to N_D. In the following paragraphs, the situation where there are less than N_D customers in the waiting list is investigated.

Based on the transition diagram in Fig. 3.17, which shows the situation where the vehicle is travelling between customers on the first trip, the differential equation is formulated as Equation (3.40).

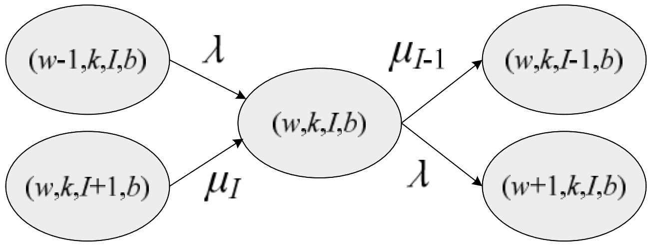


Fig. 3.17. Transition diagram of the vehicle travelling between customers on the first vehicle trip

$$\frac{d\pi'_{(w,k,I,b),t}}{dt} = \mu_I \pi'_{(w,k,I+1,b),t} + \lambda \pi'_{(w-1,k,I,b),t} - (\mu_{I-1} + \lambda)\pi'_{(w,k,I,b),t} \tag{3.40}$$

The situation of the vehicle leaving a customer's location on the first trip is illustrated in Fig. 3.18. The corresponding differential equation is formulated as Equation (3.41).

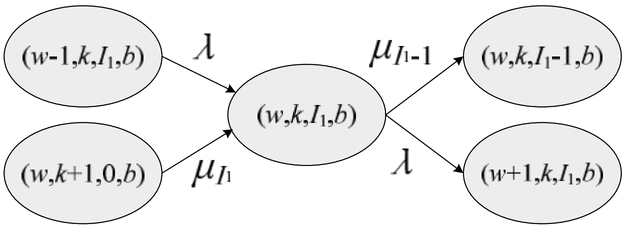


Fig. 3.18. Transition diagram of vehicle leaving customers' locations on the first vehicle trip

$$\frac{d\pi'_{(w,k,I_1,b),t}}{dt} = \mu_{I_1} \pi'_{(w,k+1,0,b),t} + \lambda \pi'_{(w-1,k,I_1,b),t} - (\mu_{I_1-1} + \lambda)\pi'_{(w,k,I_1,b),t} \tag{3.41}$$

The situation of the vehicle travelling between a customer and the warehouse on the first trip is illustrated in Fig. 3.19. The differential equation is as Equation (3.42).

$$\frac{d\pi'_{(w,0,I,b),t}}{dt} = \frac{\mu'_I}{2} \pi'_{(w,0,I+1,b),t} + \lambda \pi'_{(w-1,0,I,b),t} - \left(\frac{\mu'_{I-1}}{2} + \lambda\right)\pi'_{(w,0,I,b),t} \tag{3.42}$$

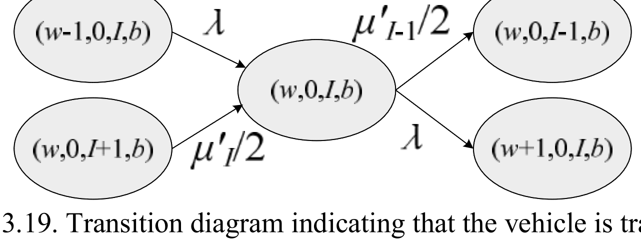


Fig. 3.19. Transition diagram indicating that the vehicle is travelling between a customer and the warehouse on the first trip.

When the vehicle returns to the warehouse and the number of customers in the waiting list is greater than or equal to N_D, the vehicle starts the second trip. The number of customers that need to be addressed on the second trip depends on the value of b.

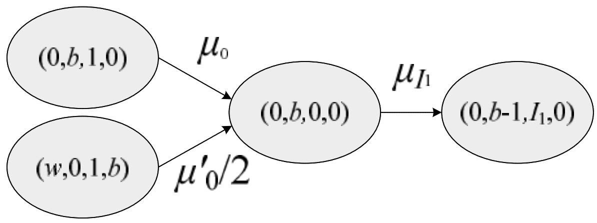
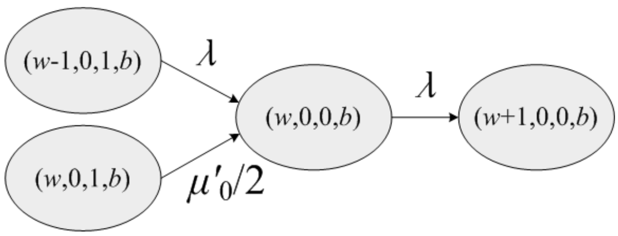


Fig. 3.20. Transition diagram in the situation that the vehicle returns to the warehouse and starts the second trip

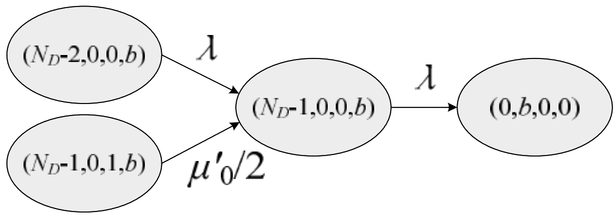
Based on the transition diagram in Fig. 3.20, the transient state differential equation is formulated as Equation (3.43).

$$\frac{d\pi'_{(0,b,0,0),t}}{dt} = \frac{\mu'_0}{2}\pi'_{(w,0,1,b),t} + \mu_0\pi'_{(0,b,1,0),t} - \mu_{I_1}\pi'_{(0,b,0,0),t} \tag{3.43}$$

When the vehicle returns to the warehouse and the number of customers in the waiting list is less than N_D, the vehicle remains idle until there are enough customers in the waiting list. Based on the transition diagram in Fig. 3.21 (a) and (b), the transient state differential equation is formulated as Equation (3.44) and Equation (3.45), respectively.



a. When $w < N_D$-1, the vehicle is idle at the warehouse.



b. When $w = N_D$-1, the vehicle starts the second trip once there is new demand.

Fig. 3.21. Transition diagram when the vehicle is idle

$$\frac{d\pi'_{(w,0,0,b),t}}{dt} = \frac{\mu'_0}{2}\pi'_{(w,0,1,b),t} + \lambda\pi'_{(w-1,0,0,b),t} - \lambda\pi'_{(w,0,0,b),t} \tag{3.44}$$

$$\frac{d\pi'_{(N_D-1,0,0,b),t}}{dt} = \frac{\mu'_0}{2}\pi'_{(N_D-1,0,1,b),t} + \lambda\pi'_{(N_D-2,0,0,b),t} - \lambda\pi'_{(0,b,0,0),t} \tag{3.45}$$

In the second trip, w and b are negligible and are both set to 0. The transition in the second trip is similar to the process illustrated in section 3.4. Based on previous design of the process, a few states becomes redundant, such as ($w,k,I,b \mid b > w$) and ($w,k,I,b \mid k > 0$ and $w > N_D$). Therefore, the state space has been significantly reduced by excluding redundant states. The construction of the intensity matrix $Q' = \{q'_{u,v}\}$ is summarized in Appendix A.1.

The solution of $\pi'_{(0,0,0),t}$ can be obtained at any time t based on Equation (3.27), subject to the initial conditions of (3.36), (3.37), (3.38) and (3.39). The service level S_T can be obtained from Equation (3.28). Please note that this model is also fit for the situation where there is no vehicle departure criterion when $N_D = 1$.

3.6. Model Validations

In order to verify the accuracy of the proposed Markov model for the parcel delivery services, a large number of experiments are conducted to compare results from both simulations and this model. The simulations are discrete-event simulations. There are three events: New Demand Event, Vehicle Departure Event and Vehicle Arrival Event. The event flow charts are illustrated in Fig. 3.22. An instance of a simulation is run for 5×10^6 units of time. After waiting for 10^5 units of time to warn up, vehicle idle time and customer waiting time are collected. This data is collected in order to evaluate performance measure of vehicle utilization, average customer waiting time and the service levels in different situations. The simulation results presented are the average values of the results obtained from 10 instances at 95% confidence level or higher associated with each scenario. Both the simulations and the calculations of the Markov model are implemented in MATLAB 7.0 on a PC with 2.33GHz CPU and 3.25GB of RAM. Solving the Markov model is computationally more efficient as it only cost several seconds, whereas the simulation takes a longer time.

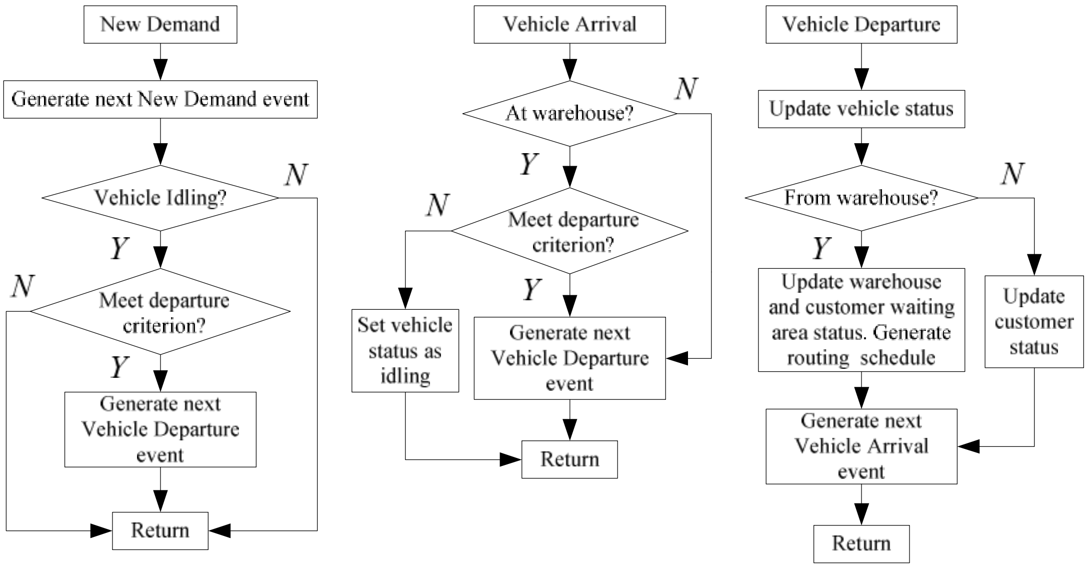


Fig. 3.22. Simulation event flow chart

3.6.1. Numerical Results for Various Demand Rates

In this section, the Markov model proposed in section 3.3 and section 3.4 will be validated. In the following experiments, the vehicle needs to serve customers who are uniformly distributed in a 100x100 square region. The customer appears in a Poisson manner with an arrival rate λ. There are four groups of simulations and Markov model calculations with $\lambda = 1/60, 1/80, 1/100, 1/120$, respectively. The vehicle starts the delivery trip from the warehouse located in the middle of the region, travels with a constant speed, and returns to the warehouse after the current trip to load products for the next trip. The loading and unloading time at the warehouse or customers' locations are not taken into consideration in both simulations and the proposed Markov model. Table 3.1 shows the average results from the simulations, results from the proposed Markov model, and their differences.

Table 3.1.
Result verification in a 100x100 square region

Demand rate		Utilization	Ave Waiting	Service Level 100	Service Level 300	Service Level 500	CPU time
Simulation	1/60	0.9669	417.1925	0.0981	0.4349	0.6876	19min
Model		0.9646	368.6309	0.0983	0.4623	0.7394	15.44s
Difference		-0.2442	-11.6401	0.1489	6.3008	7.5252	
Simulation	1/80	0.8247	153.1940	0.3887	0.8932	0.9851	20min
Model		0.8239	149.8711	0.4064	0.9065	0.9886	15.93s
Difference		-0.1000	-2.1690	4.5505	1.4895	0.3543	
Simulation	1/100	0.6956	105.8485	0.5643	0.9725	0.9989	21min
Model		0.6963	104.3074	0.6055	0.9772	0.9990	15.02s
Difference		0.1003	-1.4559	7.2940	0.4879	0.0141	
Simulation	1/120	0.5942	85.5437	0.6722	0.9911	0.9999	22min
Model		0.5973	85.3439	0.7243	0.9923	0.9999	15.27s
Difference		0.5212	-0.2334	7.7419	0.1243	0.0000	

"Utilization" represents the vehicle travelling cost; "Ave Waiting" represents the average customer waiting time for the services; "Service Level" has 3 criterions, in which the service provider promise to finish delivery within 100, 300, or 500 units of time, respectively. "Simulation" represents the results from simulation; "Model" represents the results from calculations using the proposed Markov model; "Difference" represents the differences between results of the two, calculated by 100*(Model-Simulation)/Simulation.

Table 3.1 compares the simulation and the proposed Markov model in terms of the vehicle utilization, the average customer waiting time for services and the service levels. Thee criterions are used to calculate service levels: 100, 300, 500 units of time. The service levels represent the percentage of customers' demands satisfied within, for example, 100 units of time. In the proposed Markov model, utilization is obtained from the steady state process mentioned in section 3.3, while the average customer waiting time and service levels are obtained from the customer waiting process mentioned in section 3.4.

The last column of Table 3.1 shows the CPU time comparison between simulation and the proposed model. The calculation based on the Markov model is much faster than a simulation. It is noted that the variance of the customer demand rates does not have any impact on the CPU time, since the number of states and the structure of the Markov model are the same for different customer demand rates. In contrast, the simulation time slightly increases when demand rate decreases, since system tends to generate vehicle trips visiting fewer customers and more departure and arrival events at the depot in the case of low demand rate.

The differences between results from Markov model and simulation are relatively small in terms of vehicle utilization and are all less than 0.6%. This is due to the fact that the model for the steady state process is quite accurate. Fig. 3.23 shows the steady state probability in the case of 1/100 customer demand. This figure illustrates the probability that there are 0~10 new customers acquiring parcel delivery services in the waiting list. The probability curves from the Markov model and the simulation almost overlap with each other. This evidence indicates that the Markov model approaches the vehicle utilization and transportation cost accurately.

The difference for the customer waiting process is relatively larger, due to more assumptions and approximations involved. Fig. 3.24 illustrates the cumulative distribution of the customer waiting time. Although, the difference between Markov model and simulation are visible at the beginning and the middle of the curves, the two curves are still consistent with each other. The transient state Markov Chain is able to accurately estimate the customer waiting time. Table 3.1 shows that the average customer waiting time and service levels obtained from the proposed Markov model are close to those from simulations, except in the case of 1/60 demand rate. Most of the differences are less than 4%, and the maximum difference is 7.74%, which is generated when the service level is between 60%~70%.

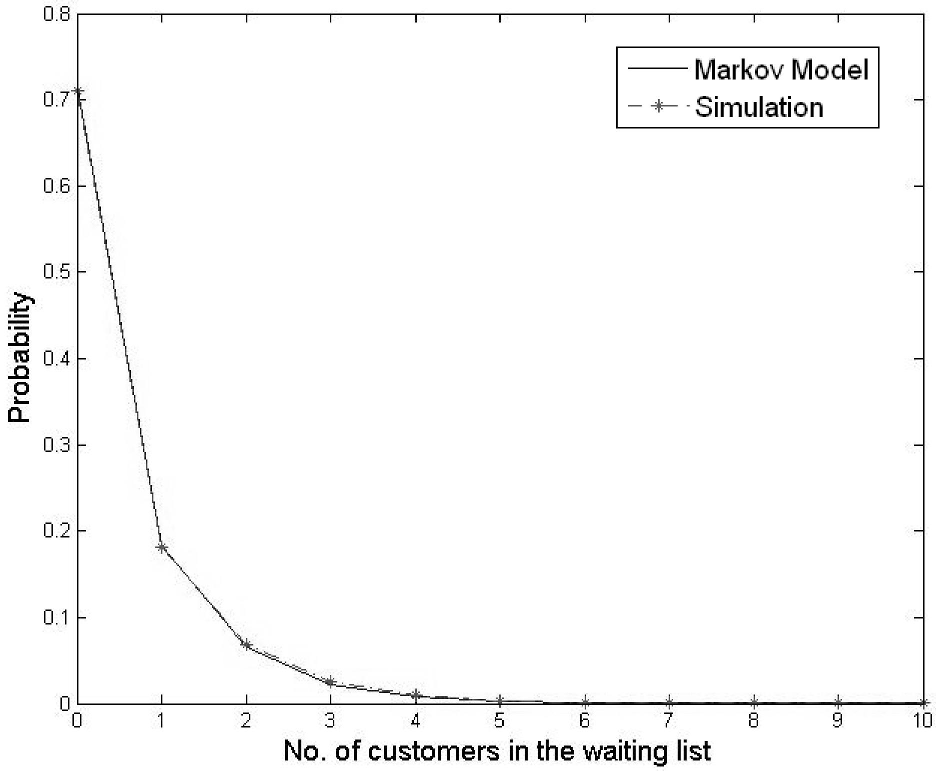


Fig. 3.23. Steady State probability of customers in waiting list (λ=1/100)

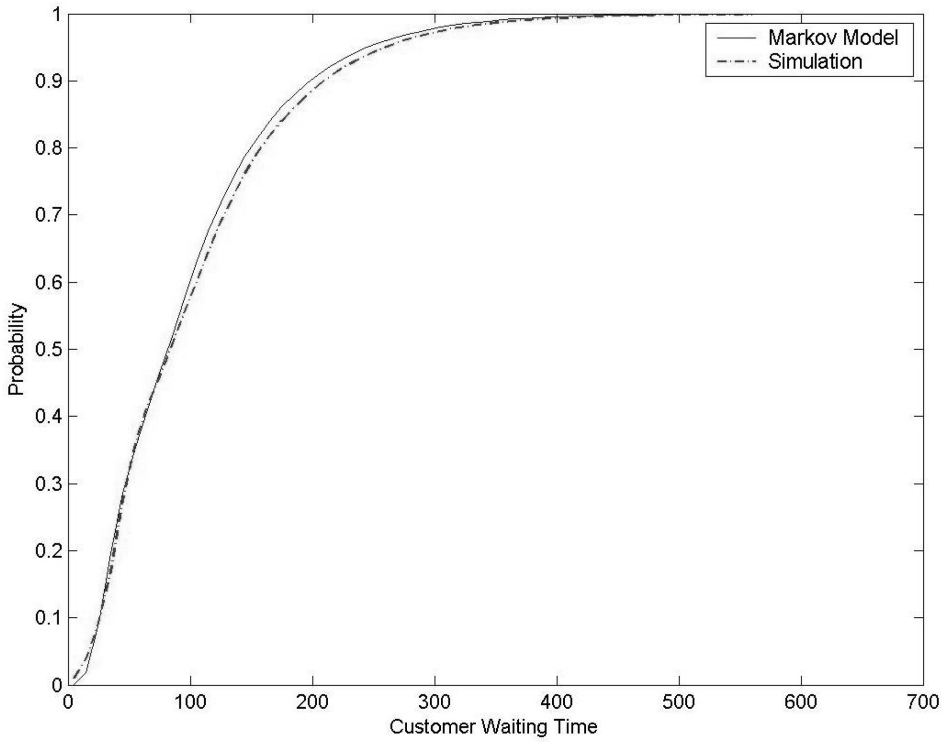


Fig. 3.24. Cumulative distribution of Customer waiting time (λ=1/100)

Even in the heavy traffic case of 1/60 customer arrival rate, the errors from customer waiting process are less than 12%. This 12% error may result from the non-sufficient state space involved in the Markov model for such high demand. In order to reduce calculation time, the maximum queue length is set to 15, which means at most 15 customers can wait for services and customers will leave when they appear and find 15 persons in the queue. The probability that the number of customers in the waiting list may exceed the queue length will increase when the customer arrival rate increases. Therefore, the average number of customers obtained from the proposed Markov model is lower than that from the simulation. The service levels from the proposed Markov model are higher than those from the simulation, since the specific customer, which is tracked to calculate the customer waiting time in the Markov model, never waits for more than 15 persons in the queue. The differences will increase when customer arrival rate increases. Extending the queue length may improve the results from the proposed Markov model. Fig. 3.25 illustrates the variance of calculation errors and CPU time by increasing the queue length. When the queue length increases, the error decreases with a significant increase of the CPU time. For example, when the queue length of the proposed Markov model is extended to 22, it takes 2.5 times of calculation effort to obtain 5% error in terms of the average customer waiting time and less than 5% errors in terms of service levels.

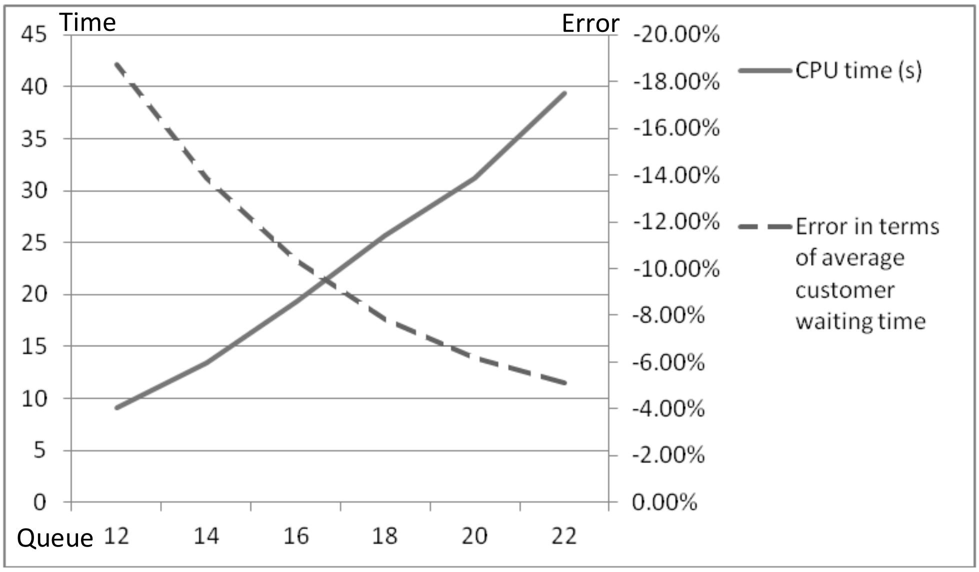


Fig. 3.25. The calculation errors and CPU time vary in the setting of queue length

In Table 3.1, when the demand rate increases, the vehicle utilization increases, as more customers need service in the same period of time. Meanwhile, the average customer waiting time increases, and service levels decrease as well. Customer needs to wait for longer time, since the facilities will be slightly insufficient in heavy traffic situations. It can also be seen that almost all of the differences which arise between results from the proposed Markov model and the simulation increases with the increase of demand rate. For example, the difference in terms of average customer waiting time in the case with 1/80 demand rate is almost 2% higher than that in the case with 1/120 demand rate. This evidence implicates that the proposed Markov model is more accurate in light traffic condition than heavy traffic condition, due to the setting of the queue length in the proposed Markov model.

3.6.2. Numerical Results for Vehicle Departure Strategies

In this section, the Markov model proposed in section 3.5 will be validated. In the second set of experiments, every condition is the same as the previous experiments except that the customer arrival rate is fixed at $\lambda = 1/80$ and results are compared among different vehicle departure criterions, in which the vehicle is allowed to start a new trip when there are at least 1~6 customers in the waiting list, respectively. The results are shown in Table 3.2.

The calculation based on the Markov model is much faster than a simulation in the CPU time comparison. It is noted that the increase of the departure criterion complicates the calculation in the customer waiting process, hence the CPU time of the proposed model increases. In contrast, the simulation time decreases when the departure criterion increases, since system tends to generate vehicle trips visiting more customers and fewer departure and arrival events at the depot in the case of higher departure criterion.

Again, Table 3.2 shows that the results obtained from the proposed Markov model are close to those from simulations. Especially for the vehicle utilization, the differences are within 1%. The maximum difference in results generated from the customer waiting process is 4.55%. Most of the differences are less than 2.5%. It can be noted that there are no results for service levels within 100 units of time from simulation in the cases when the vehicle departure criterion is from three to six. Since there are few demands which can be fulfilled within 100 units of time, there is no enough evidence in a simulation to conclude the service levels. Furthermore, the service levels in these cases are small, which are less than 0.05, and the calculation error indicated in inequation (3.30) would be significant.

In Table 3.2, when the vehicle departure criterion increases, the vehicle utilization will slightly decrease, as the vehicle may stay at the warehouse for longer time to wait for enough demands to be generated. Meanwhile, the average customers waiting time increases significantly, and the differences decrease from 2.17% to 0.68%. Furthermore, the service levels decrease rapidly when the vehicle departure criterion increases, since customers have to spend more time waiting for enough demands to warrant a delivery trip. In the meantime, the differences between the Markov model and simulation increase.

Table 3.2.
Result verification with customer demand rate 1/80

Departure Criterion		Utilization	Ave Waiting	Service Level			CPU time
				100	300	500	
Simulation	1	0.8247	153.1940	0.3887	0.8932	0.9851	21min
Model		0.8239	149.8711	0.4064	0.9065	0.9886	15.24s
Difference		-0.1000	-2.1690	4.5505	1.4895	0.3543	
Simulation	2	0.7649	188.2145	0.2113	0.8578	0.9823	15min
Model		0.7700	185.6798	0.2121	0.8679	0.9841	15.17s
Difference		0.6684	-1.3467	0.3769	1.1815	0.1832	
Simulation	3	0.7367	237.2031		0.7626	0.9690	13min
Model		0.7392	234.6209	0.0489	0.7738	0.9719	17.38s
Difference		0.3514	-1.0886		1.4597	0.2988	
Simulation	4	0.7187	291.6203		0.6024	0.9385	12min
Model		0.7208	288.8803	0.0056	0.6124	0.9425	18.69s
Difference		0.2883	-0.9396		1.6597	0.4223	
Simulation	5	0.7085	349.6218		0.4036	0.8778	12min
Model		0.7085	347.2129	0.0005	0.4126	0.8844	20.75s
Difference		0.0001	-0.6890		2.2158	0.7526	
Simulation	6	0.6985	409.4676		0.2234	0.7793	11min
Model		0.6998	406.6914	0.00003	0.2242	0.7886	24.55s
Difference		0.1911	-0.6780		0.3545	1.1990	

The symbols and abbreviations are the same as Table 3.1.

In summary, the results generated from the Markov model are close to the results from simulations. The Markov model is computationally more efficient than simulations. Therefore, the Markov model can replace the simulations in order to quickly and accurately estimate performance measures for parcel transportation services.

4. Extension and Modification of the Markov Model

4.1. Overview

In Chapter 3, parcel delivery services were analyzed in a Markov model. However, several assumptions were made, such as only one vehicle serving the entire region, the unlimited capacity of the vehicle, and the FCFS vehicle routing strategy. In this chapter, these assumptions will be relaxed, and the corresponding modifications of the Markov model will be applied in order to adapt these new constraints and features. In section 4.2, services with limited vehicle capacity constraints are analyzed. In section 4.3, services with more than one vehicle delivering products in the same region is examined. The minimum number of vehicles needed will be decided after investigating parcel delivery services with multiple vehicles. In section 4.4, the difference between delivery services and pickup services is discussed. The Markov model constructed in chapter 3 is held valid for delivery services, and another Markov model is built for pickup services in this section. In section 4.5, several vehicle routing strategies, such as insertion based algorithms and Branch-and-Bound based algorithms are investigated. Numerical results demonstrate that the extended Markov models are capable of providing fast and reliable performance measures estimations in the various situations mentioned above.

4.2. Issue of Vehicle Capacity

In previous discussions, it has been assumed that the vehicle has infinite capacity, since the sizes of products ordered by customers are relatively small. This is the case where the vehicle can visit as many customers as possible in one trip. In the interest of efficiency, the vehicle carries all the products ordered by customers in the waiting list when it starts a new trip.

However, it may not be realistic to make the assumption of infinite capacity when larger sizes of parcels are taken into consideration. In order to take the first step in studying the Markov model for the capacity issue, it is assumed that all parcels are the same size. In this case, a parcel to be delivered to a customer occupies one unit space of the vehicle capacity, and the limit of vehicle capacity C can be considered as the maximum number of customers allowed in a trip. Hence, the vehicle carries parcels for the first C customers in the waiting list based on the FCFS rule. Any customers after the first C have to be queued in the waiting list until the following trip. Based on the calculation in sections 3.3 and 3.4, several formulas need to be modified in order to adapt the vehicle capacity constraint.

4.2.1. Model Modification for the Capacity Issue

When $w \leq C$, Fig. 3.6 and Equation (3.12) are still valid. However, when $w > C$, the vehicle starts the new trip with only C customers, and the rest remains in the waiting list. The state transition is shown in Fig. 4.1.

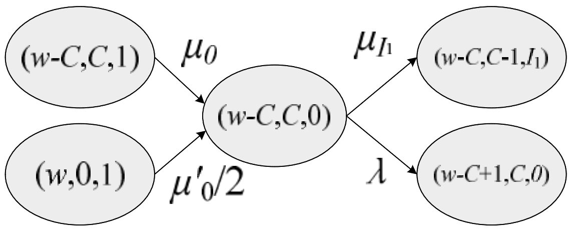


Fig. 4.1. Transition diagram of $w>C$ customers in the waiting list when the vehicle starts a new trip

The balance equation for this transition diagram is as follows.

$$\frac{\mu_0'}{2}\pi_{(w,0,1)} + \mu_0\pi_{(w-C,C,1)} = (\lambda + \mu_{I_1})\pi_{(w-C,C,0)} \tag{4.1}$$

Since the balance equations have been modified, several elements in the intensity matrix Q are modified as follows.

$$q_{(w,0,1),(w-C,C,0)} = \mu_0'/2\text{ , } w = C+1, C+2,\cdots$$

In terms of the customer waiting process, the service for the specific customer may not be scheduled on the second vehicle trip if more than C customers are waiting. However, the process can still be divided into two steps. The first step is vehicle trips for customers before the specific one; the second step includes the service for the specific customer. The customer waiting process in section 3.5 can be modified as follows.

1) The remainder of w-1 divided by C is greater or equal to N_D-1 when the vehicle is travelling between customers' locations.

   In this situation, demand after the specific customer will not affect the rest of the process, since the number of customers in the waiting list is enough to trigger the vehicle trip for the specific customer. It is not necessary to track the number of customers in the waiting list. Furthermore, the reconstructed process is similar to the one in section 3.5.

   Initially, $w = N_D$ and $b = 1$ for all situations in this category, where b is only a notation instead of a real position of the specific customer. The initial state probabilities are formulated as follows.

$$\pi'_{(N_D+Cj,k,I,1),0} = \sum_{i \ge N_D-1}^{\min(C-1,k-1)} \pi_{(i+Cj,k-i,I)}\text{ , } k = N_D, N_D+1,\cdots 2C; I, j = 0,1,\cdots \tag{4.2}$$

2) The remainder of w-1 divided by C is less than N_D-1 when the vehicle is travelling between customers.

   In this situation, the number of customers in the waiting list should be accumulated until N_D+Cj ($j = 0,1,2,\cdots$) in order to trigger a vehicle trip for the specific customer. The number of customers w in the waiting list will increase by one when demand is generated. Once w reaches nearest N_D+Cj, new demand will not be considered, as it will not affect the following process.

b records the position of the specific customer on the final trip, and b-1 is the remainder of w-1 divided by C. b is fixed in the process before the last vehicle trip. w keeps increasing if new demands appear until $w = N_D + Cj$. The probabilities for the initial state are formulated as follows.

$$\pi'_{(w,k,I,b),0} = \pi_{(w-1,k,I)}, \quad w = b + Cj; k = 1, 2, \cdots; b = 1, 2, \cdots, N_D - 1; I, j = 0, 1, \cdots \qquad (4.3)$$

3) The remainder of w-1 divided by C is greater or equal to N_D-1 when the vehicle travels between a customer and the warehouse.

In this situation, the vehicle is heading to the warehouse, and there are an adequate number of customers in the waiting list to trigger a final trip for the specific customer. No new demand will be taken into consideration in the process. $b-1$ is the remainder of $w-1$ divided by C. The initial state probabilities are obtained as follows.

$$\pi'_{(w,0,I,b),0} = \pi_{(w-1,0,I)}, \quad w = b + Cj; b = N_D, N_D + 1, \cdots C - 1; I, j = 0, 1, \cdots \qquad (4.4)$$

4) The remainder of w-1 divided by C is less than N_D-1, when the vehicle is travelling between a customer and the warehouse.

Similar to situation 2), b is the position of the specific customer on the last vehicle trip, and b-1 is the remainder of w-1 divided by C. w continues increasing if new demand appears. Once there are enough customers in the waiting list, no new demands will be considered in the process. The probabilities of the initial state are obtained as Equation (4.5).

$$\pi'_{(w,0,I,b),0} = \pi_{(w-1,0,I)}, \quad w = b + Cj; b = 1, 2, \cdots, N_D - 1; I, j = 0, 1, \cdots \qquad (4.5)$$

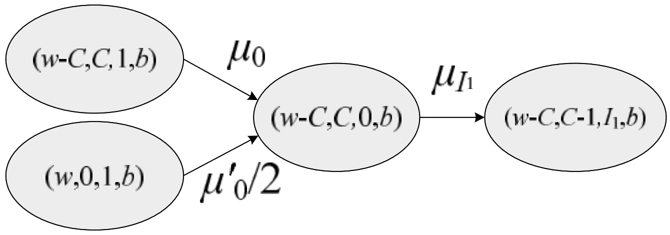
When the number of customers in the waiting list is less than C, the customer waiting processes are similar to those in section 3.4 and 3.5. Even if the number of customers in the waiting list is greater than C, the vehicle travelling can still be described as previously mentioned model. In the case when the remainder of w-1 divided by C is greater or equal to N_D-1, the transition diagrams of the vehicle travelling are illustrated in Fig. 3.9, 3.10 and 3.12. The corresponding differential equations are formulated in Equations (3.21), (3.22) and (3.24). In the case when the remainder of w-1divided by C is less than N_D-1, the transition diagrams of vehicle travelling are illustrated in Fig. 3.17, 3.18 and 3.19. The corresponding differential equations can be seen in Equations (3.40), (3.41) and (3.42). The differences in the process happen when the vehicle finishes a trip and return to the warehouse.

When the vehicle returns to the warehouse and the number of customers in the waiting list is less than C, the process is similar to the cases illustrated in Fig. 3.11 and 3.21. The corresponding differential equations can be seen in Equations (3.23), (3.44) and (3.45), respectively. When the vehicle returns to the warehouse and the number of customers in the waiting list is greater than C, the vehicle starts its next trip, but the specific customer is still queued in the waiting list.

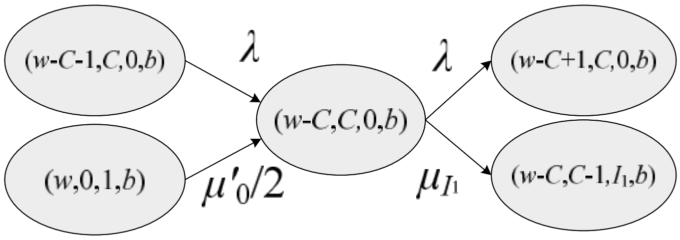
Based on the transition diagram in Fig. 4.2, the transient state differential equation is formulated as Equations (4.6) and (4.7), respectively.

$$\frac{d\pi'_{(w-C,C,0,b),t}}{dt} = \frac{\mu'_0}{2}\pi'_{(w,0,1,b),t} + \mu_0\pi'_{(w-C,C,1,b),t} - \mu_{I_1}\pi'_{(w-C,C-1,I_1,b),t} \tag{4.6}$$

$$\frac{d\pi'_{(w-C,C,0,b),t}}{dt} = \frac{\mu'_0}{2}\pi'_{(w,0,1,b),t} + \lambda\pi'_{(w-C-1,C,0,b),t} - (\lambda + \mu_{I_1})\pi'_{(w-C,C,0,b),t} \tag{4.7}$$



a. the remainder of w-1 divided by C is greater or equal to N_D-1



b. the remainder of w-1 divided by C is less than N_D-1

Fig. 4.2. Transition Diagram when the vehicle returns to the warehouse and starts the next trip

w and b are negligible and are both set to zero, when the vehicle sets out on a trip to the specific customer. The transition is similar to the process illustrated in section 3.4. The construction of the intensity matrix $Q' = \{q'_{u,v}\}$ is summarized in Appendix A.2.

The solutions for $\pi'_{(0,0,0),t}$ can then be obtained based on Equation (3.27). The vehicle utilization and transportation cost are also obtained by Equation (3.34) and Equation (3.35). The service level S_T can be calculated from Equation (3.28).

4.2.2. Model Validations and Result Discussions

The following experiments are carried out in order to verify the accuracy of the proposed Markov model for the capacity issue in parcel delivery services. In the experiments, the vehicle delivers products to customers who are uniformly distributed within a 100x100 square region. Customers appear in a Poisson manner with an arrival rate $\lambda = 1/80$. Four groups of simulations and Markov model calculations with vehicle capacity C equal to 3, 5, 8 and 12 are investigated, respectively. The vehicle starts the delivery trip from the warehouse located in the middle of the region once there is pending demand, and travels with a constant speed. On each trip, the vehicle can only visit at most 3, 5, 8, or 12 customers, respectively. The loading and unloading time at the warehouse or customers' locations are not taken into consideration in both simulations and the proposed Markov model. Table 4.1 shows the average results from simulations, results from the proposed Markov model, and their differences.

Table 4.1.
Result verification with customer demand rate 1/80

Capacity		Utilization	Ave Waiting	Service Level			CPU time
				100	300	500	
Simulation	3	0.8307	168.8295	0.3724	0.8544	0.9663	21min
Model		0.8386	162.4710	0.3730	0.8712	0.9775	8.88s
Difference		0.9534	-3.7662	0.1468	1.9663	1.1591	
Simulation	5	0.8255	154.1252	0.3869	0.8922	0.9826	21min
Model		0.8319	152.2400	0.3881	0.8982	0.9846	8.12s
Difference		0.7749	-1.2231	0.3132	0.6719	0.2060	
Simulation	8	0.8258	152.5175	0.3856	0.8956	0.9875	21min
Model		0.8310	150.0725	0.3902	0.9037	0.9875	10.06s
Difference		0.6265	-1.6031	1.1786	0.9053	0.0000	
Simulation	12	0.8261	151.5263	0.3858	0.8999	0.9878	22min
Model		0.8309	149.8005	0.3903	0.9040	0.9880	10.77s
Difference		0.5809	-1.1389	1.1644	0.4554	0.0186	

"Utilization" represents the vehicle travelling cost; "Ave Waiting" represents the average customer waiting time for the services; "Service Level" has 3 criterions, in which the service provider promise to finish delivery within 100, 300, or 500 units of time, respectively. "Simulation" represents the results from simulation; "Model" represents the results from calculations using the proposed Markov model; "Difference" represents the differences between the two, calculated by 100*(Model-Simulation)/Simulation.

Table 4.1 compares the simulation and the proposed Markov model in terms of vehicle utilization, average customer waiting time and service levels. When the vehicle capacity increases, the vehicle utilization will decrease as the vehicle effectively combines the services for more customers in a single trip. As a result, both the travelling costs and the average customer waiting time are reduced and service levels increase. Customer tends to wait for shorter period of time, since vehicle routing efficiencies are achieved. It can be noted that the results of the vehicle with capacity 12 are close to the results of vehicle with infinite capacity in Table 3.1, since there is rare chance that more than 12 customers will be queued in the waiting list when the demand rate is 1/80.

In the comparison of the CPU time between simulation and the proposed model, the calculation based on the Markov model is much faster than a simulation. When the capacity of the vehicle is small, the maximum number of stops in a vehicle trip is limited and the number of states is small. However, the calculation in the customer waiting process is complicated, since the vehicle may start the trip including the service for the specific customer after several trips. This may be the reason why the CPU time increases after the initial decrease. In contrast, the variance of the vehicle capacity does not have significant impact on the CPU time of the simulation.

The differences between results from the Markov model and the simulation are minor in terms of vehicle utilization and service levels. Most of the differences are less than 2%, and the maximum is 3.77%. It can be also noted that almost all of the differences decrease when vehicle capacity increases. For example, in terms of vehicle utilization, the difference for a vehicle with capacity 3 is almost 0.4% higher than the case of capacity 12. In summary, the Markov model is able to accurately estimate the performance of the delivery services using vehicles with different capacities.

4.2.3. Case Study of Vehicle Selection

In this set of experiments, the proposed model is used to decide which vehicle to utilize from a pool of different types of vehicles. Larger vehicles have larger capacities but may consume more petrol per unit time. The characteristics of vehicles are listed in Table 4.2. In the experiments, the vehicle delivers products to customers who are uniformly distributed in a 100x100 square region. Customers appear in a Poisson manner with an inter-arrival rate $\lambda = 1/80$. The service provider promises that customer demand will be fulfilled within 300 units of time; otherwise the customer will acquire $50 compensation for each overdue service. It is assumed that the vehicle travels with a constant speed with a petrol cost φ_1 per unit time. The estimated cost per unit time, which includes the cost of the vehicle travelling and the penalty of overdue services, is calculated by Equation (4.8), and the results are also shown in Table 4.2.

$$\text{Cost} = \varphi_1 U + 50(1 - S_T)\lambda \quad (4.8)$$

Table 4.2.
Parameters of vehicles and total costs

C	2	3	4	5	6	7	8
φ_1	0.22	0.23	0.24	0.25	0.26	0.27	0.28
Cost	0.2957	0.2734	0.2694	0.2716	0.2774	0.2849	0.2928

"C" is the capacity of the vehicle; "φ_1" is the petrol cost per unit time when the vehicle is travelling; "Cost" is calculated by Equation (4.8).

Table 4.1 has concluded that as the vehicle capacity increases, the vehicle tends to generate more efficient schedules for services and service levels increase as well. However, the trade-off between a larger vehicle with higher efficiency and higher consumption of petrol per unit time must be taken into consideration. Table 4.2 shows that the total transportation cost first decreases, then reaches a minimum value, and finally increases as the vehicle capacity increases. If the appropriate vehicle is selected based on the minimum total cost estimated, the vehicle with capacity 4 will be choosen. It would expend $0.2694 on travelling per unit time.

4.3. Issue of Multiple Vehicles and Vehicle Management

Previously it was assumed that only one vehicle was running in a selected service region. In this section, this assumption can be relaxed. It may not be practical that a fixed sub-region will be assigned to a certain vehicle. The service regions assigned to different vehicles may overlap with each other. That is the reason why most papers on routing strategies for the DVRP, in past research, take multiple vehicles into consideration.

The research done in this thesis is the first time that the Markov model has been studied for multiple vehicles in parcel delivery services. If N_V vehicles are available to deliver products to customers within the same region, it is assumed that all vehicles are identical and have equal opportunities to serve customers. For example, a new customer with a delivery request appears at a certain time. If there is no vehicle idle at the warehouse, the new customer has to wait until a vehicle finishes its trip. This new delivery request will be handled by the first vehicle that returns, if the departure criterion is met. If there are N_V vehicles idle at the warehouse, each one has $1/N_V$ chance to start a new trip and to fulfil the new request.

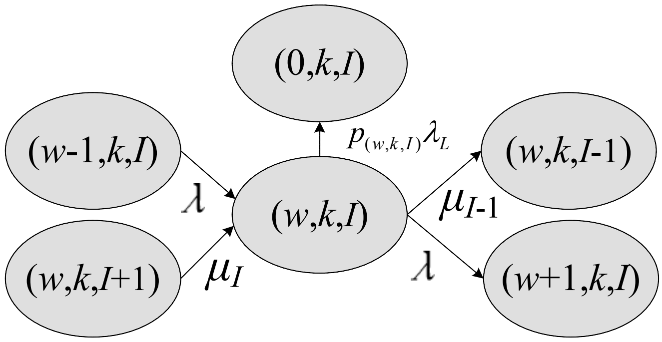
4.3.1. Model Modification for the Multiple Vehicles Issue

As previously mentioned, the system state is represented by (w,k,I), where w is the status of the waiting list, k and I represent the status of the vehicle. Now, there is more than one vehicle, and two additional variables are needed to represent the status of each vehicle. For example, if there are three vehicles in the same service region, the system state should be $(w,k_1,I_1,k_2,I_2,k_3,I_3)$. In this case, there are ten million ($10^{2\times3+1}$) states if each parameter has 10 different values, and the intensity matrix is $10^7\times10^7$. Hence, the state space is too large to obtain a solution for this problem. Approximation methods have to be used to simplify the Markov process.

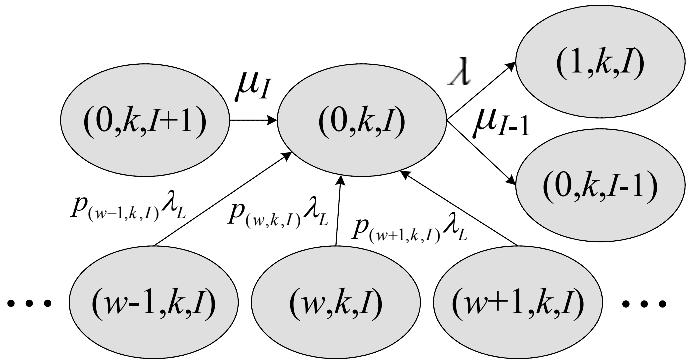
Since all vehicles are identical and have equal chances to serve customers, it is unnecessary to differentiate them. On average, the number of customers' requests fulfilled and the vehicle idle time will be the same no matter which vehicle is observed. Therefore, the problem can be analyzed by focusing on only one of the vehicles. The system state is still defined as (w,k,I), where k and I are the status of the observed vehicle. Meanwhile, the state transitions due to the operation of other vehicles have to be specified as follows.

It is assumed that each vehicle has to spend time t_L to prepare for the next new trip. t_L is a random variable independent of number of customers visited on the following trip, and it follows an exponential distribution with λ_L=1. Due to this assumption, the states of (w,0,0) are valid, since each vehicle has to spend some time at (w,0,0) to prepare for the following trip.

Based on previous discussion, the observed vehicle travels on the road and visits customers one after another according to the schedule. Once a new customer appears, the number of customers in the waiting list increases by one. However, there is a chance that the number of customers in the waiting list will suddenly become zero, even though the observed vehicle has not yet returned to the warehouse. This happens when another vehicle returns to the warehouse and plans a new trip for all customers in the waiting list. To clarify, the case with departure criterion of one is used to explain this process. Fig. 4.3 shows the transition diagram when the observed vehicle en route.



a. *w* customers in the waiting list



b. no customer in waiting list

Fig. 4.3. Transition diagram of the vehicle travelling

The balance equations for these transition diagrams are formulated as Equation (4.9) and Equation (4.10), respectively.

$$\mu_I \pi_{(w,k,I+1)} + \lambda \pi_{(w-1,k,I)} = (\mu_{I-1} + \lambda + p_{(w,k,I)}\lambda_L)\pi_{(w,k,I)} \tag{4.9}$$

$$\mu_I \pi_{(w,k,I+1)} + \sum_{w \geq 1} p_{(w,k,I)}\lambda_L \pi_{(w,k,I)} = (\mu_{I-1} + \lambda)\pi_{(w,k,I)} \tag{4.10}$$

In Equation (4.9) and Equation (4.10), $p_{(w,k,I)}$ is the probability that another vehicle starts a new trip with *w* customers in the waiting list when the observed vehicle is not at the warehouse (on state (*w*,*k*,*I*)). $p_{(w,k,I)}$ is unknown, but can be calculated based on the steady state probability.

$$\begin{aligned} p_{(w,k,I)} &= P\{\text{at least one of other vehicles at warehouse} \mid (w,k,I)\} \\ &= 1 - P\{\text{no other vehicles at warehouse} \mid (w,k,I)\} \\ &= 1 - (1 - P\{\text{the other vehicle at warehouse} \mid (w,k,I)\})^{N_V - 1} \end{aligned} \tag{4.11}$$

It is assumed that the probability of the other vehicle at the warehouse p_w depends only on the number of customers in the waiting list, and is independent of the status of the observed vehicle, which is denoted as follows.

$$\begin{aligned} p_w &= P\{\text{the other vehicle at warehouse} \mid w\} \\ &= P\{\text{the other vehicle at warehouse} \mid (w,k,I)\} \end{aligned} \quad \text{for } \forall k, I$$

Based on the steady state probability, p_w can be calculated by the following equation.

$$p_w = \frac{\pi_{(w,0,0)}}{\sum_{k,I} \pi_{(w,k,I)}} \tag{4.12}$$

Accordingly, Equation (4.11) can be revised to Equation (4.13)

$$p_{(w,k,I)} = 1 - (1 - p_w)^{N_V - 1}, \text{ for } \forall (w,k,I) \neq (w,0,0) \tag{4.13}$$

When the observed vehicle returns to the warehouse and there are enough customers in the waiting list to trigger the next vehicle trip, all vehicles idle at the warehouse have an equal opportunity to commence the trip. If this new trip is handled by other vehicles, the system state transitions from (w,0,0) to (0,0,0), and the observed vehicle is still idle at the warehouse. If the observed vehicle starts the new trip, the system state transitions to (0,w,0), accordingly. Fig. 4.4 illustrates the transition diagram when the observed vehicle is idle at the warehouse.

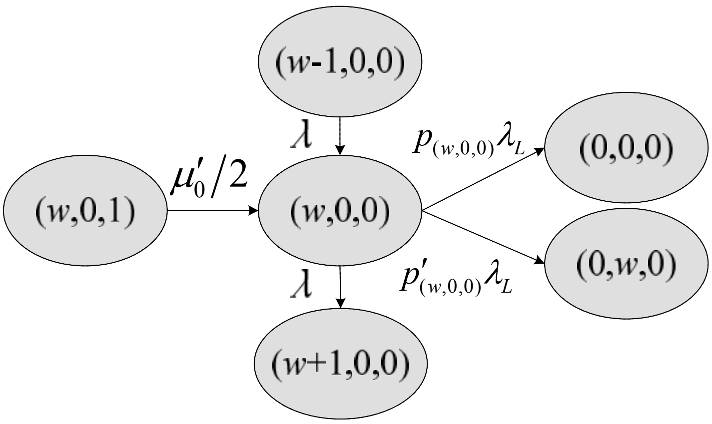


Fig. 4.4. Transition diagram of vehicle idle at the warehouse (w>0)

The balance equation of the transition diagram is as follows.

$$\frac{\mu_0'}{2}\pi_{(w,0,1)} + \lambda \pi_{(w-1,0,0)} = (\lambda + p_{(w,0,0)}\lambda_L + p'_{(w,0,0)}\lambda_L)\pi_{(w,0,0)} \tag{4.14}$$

When the observed vehicle returns to the warehouse and there are not enough customers in the waiting list to trigger the next trip, all vehicles have to wait. The transition diagram of this situation is illustrated in Fig. 4.5.

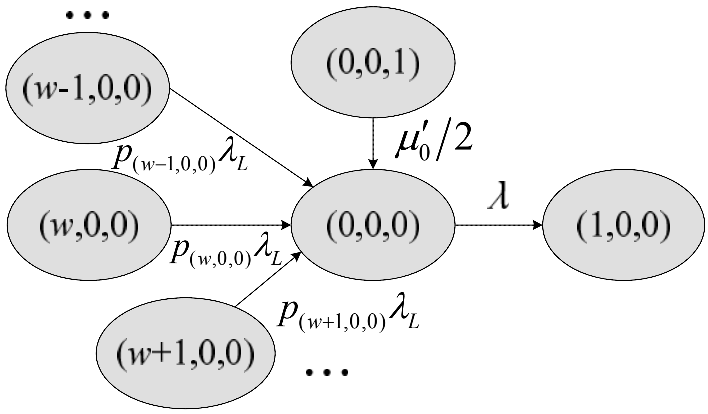


Fig. 4.5. Transition diagram of vehicle idle at the warehouse with no customers in the waiting list

Based on the transition diagram in Fig. 4.5, the balance equation is shown as Equation (4.15).

$$\frac{\mu_0'}{2}\pi_{(0,0,1)} + \sum_{w\geq 1} p_{(w,0,0)}\lambda_L \pi_{(w,0,0)} = \lambda \pi_{(0,0,0)} \tag{4.15}$$

In Equation (4.14) and Equation (4.15), $p_{(w,0,0)}$ is the probability that another vehicle starts a new trip when there are *w* customers in the waiting list with the observed vehicle at the warehouse, and $p'_{(w,0,0)}$ is the probability that the observed vehicle start a new trip itself. The values of these two variables are calculated as Equation (4.16) and Equation (4.17), respectively.

$$\begin{aligned} p_{(w,0,0)} &= \sum_{i=1}^{N_V-1} \frac{i}{i+1} P\{i \text{ of other vehicles at warehouse} \mid (w,0,0)\} \\ &= \sum_{i=1}^{N_V-1} \frac{i}{i+1} C_{N_V-1}^{i} p_w^i (1-p_w)^{N_V-1-i} \end{aligned} \tag{4.16}$$

$$
\begin{aligned}
p'_{(w,0,0)} &= 1 - p_{(w,0,0)} \\
&= \sum_{i=0}^{N_V-1} \frac{1}{i+1} P\{i \text{ of other vehicles at warehouse} \mid (w,0,0)\} \\
&= \sum_{i=0}^{N_V-1} \frac{1}{i+1} C_{N_V-1}^{i} p_w^i (1-p_w)^{N_V-1-i}
\end{aligned}
\tag{4.17}
$$

In Equation (4.16) and Equation (4.17), $C_{N_V-1}^{i}$ is the N_V-1 combination from *i* elements.

$$
C_{N_V-1}^{i} = \frac{(N_V-1)!}{i!(N_V-1-i)!}
$$

Due to modification of the balance equations, the elements in the intensity matrix Q are modified as follows.

$q_{(w,0,1),(w,0,0)} = \mu'_0/2, \ w = 0,1,\cdots$

$q_{(w,0,0),(w+1,0,0)} = \lambda, \ w = 0,1,\cdots$

$q_{(w,k,I),(0,k,I)} = \lambda_L \left[1-(1-p_w)^{N_V-1}\right], \ w = 1,2,\cdots; \forall (w,k,I) \neq (w,0\ 0)$

$q_{(w,0,0),(0,0,0)} = \lambda_L \left[\sum_{i=1}^{N_V-1} \frac{i}{i+1} C_{N_V-1}^{i} p_w^i (1-p_w)^{N_V-1-i}\right], \ w = 1,2,\cdots$

$q_{(w,0,0),(0,w,0)} = \lambda_L \left[\sum_{i=0}^{N_V-1} \frac{1}{i+1} C_{N_V-1}^{i} p_w^i (1-p_w)^{N_V-1-i}\right], \ w = 1,2,\cdots$

The diagonal elements of the intensity matrix are calculated as follows.

$q_{(w,k,I),(w,k,I)} = -\sum_{v \neq (w,k,I)} q_{(w,k,I),v}, \ \forall (w,k,I)$

An iteration is applied to calculate p_w and the steady state probability π.

1) Initialization. $p_w = 0.1$
2) Calculate $p_{(w,k,I)}$, $p_{(w,0,0)}$ and $p'_{(w,0,0)}$ based on Equations (4.13), (4.16) and (4.17).
3) Solve the model and obtain steady state probability π by Equation (3.16)

4) If $\left| p_w - \frac{\pi_{(w,0,0)}}{\sum_{k,I} \pi_{(w,k,I)}} \right| > \varepsilon_p$ (ε_p is a small value which indicates the acceptable error),

let $p_w = \frac{\pi_{(w,0,0)}}{\sum_{k,I} \pi_{(w,k,I)}}$ and repeat steps in 2), 3) and 4) until $\left| p_w - \frac{\pi_{(w,0,0)}}{\sum_{k,I} \pi_{(w,k,I)}} \right| \le \varepsilon_p$.

In the customer waiting process, the demand of the specific customer can be fulfilled by any vehicle. Therefore, it is not necessary to keep track of a certain vehicle, since the trip for the specific customer may be completed before the observed vehicle is freed up. As usual, the customer waiting process is divided into two steps. The first step is the specific customer waiting for his/her delivery to be scheduled. The second step consists of the vehicle trip including the specific customer. (w,k,I) is used to represent the system state. In the first step, the number of customers in the waiting list w continues to grow as new demand increases. k records the position of the specific customer in the waiting list, and the customer will be the k^{th} person served in the following trip. I is artificially set to 0. In the second step, one of the vehicles will take care of the delivery for the specific customer. w becomes 0, since no new demand will be taken into account. System transitions from state $(0,k,0)$ to $(0,0,0)$ after several transition steps as vehicle travelling on the road, which is the same as in section 3.4.

Since the status of the observed vehicle is not utilized, the steady state probabilities for different vehicle statuses are summed up to generate initial state probabilities for the customer waiting process Equation (4.18). Initially, the position of the specific customer in the waiting list is the same as the total number of customers in the waiting list, since the process starts at the time the specific customer appears.

$$\pi'_{(w,w,0),0} = \sum_{k,I} \pi_{(w-1,k,I)} \;, \quad w = 1, 2, \cdots \tag{4.18}$$

Since the probability that a vehicle will start a trip including the specific customer depends on the number of customers in the waiting list *w*, the system has to take new demand into consideration. Fig. 4.6 shows the transition diagram of the specific customer in the waiting list, when the system state is (*w*,*k*,0).

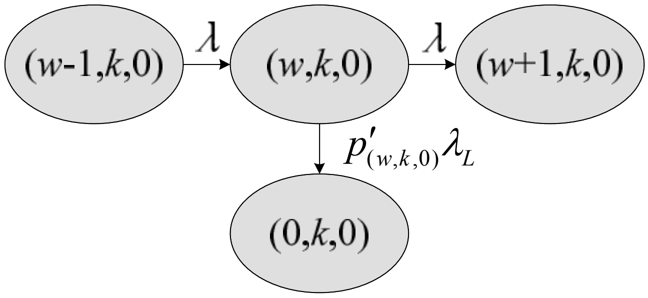


Fig. 4.6. Transition diagram of the specific customer queued in the waiting list

Based on the transition diagram, the transient state differential equation is formulated as Equation (4.19).

$$\frac{d\pi'_{(w,k,0),t}}{dt} = \lambda\pi'_{(w-1,k,0),t} - (\lambda + p'_{(w,k,0)}\lambda_L)\pi'_{(w,k,0),t} \tag{4.19}$$

In any of the states (*w*,*k*,0), there is a chance that one of the vehicles may return to the warehouse and start the following trip including the specific customer. The situation is illustrated in Fig. 4.7 and Equation (4.20).

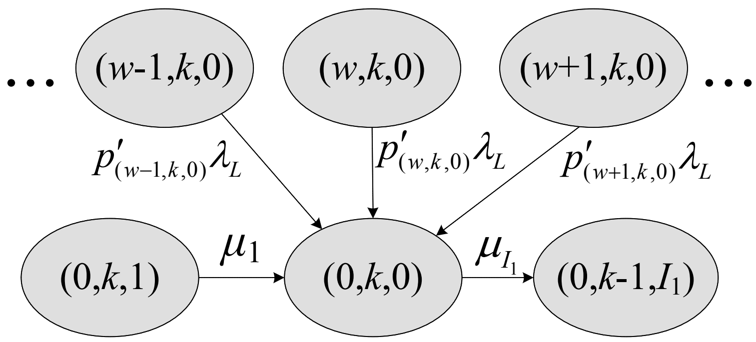


Fig. 4.7. Transition diagram of a vehicle starting a trip including the specific customer

$$\frac{d\pi'_{(0,k,0),t}}{dt} = \sum_{w} p'_{(w,k,0)} \lambda_L \pi'_{(w,k,0),t} + \mu_1 \pi'_{(0,k,1),t} - \mu_{I_1} \pi'_{(0,k,0),t} \qquad (4.20)$$

In Equation (4.18) and Equation (4.19), $p'_{(w,k,0)}$ is the probability that one of the vehicles stays in the warehouse at the time when there are w customers in the waiting list. The parameters can be calculated based on steady state probabilities.

$$\begin{aligned} p'_{(w,k,0)} &= P\{\text{at least one vehicle at warehouse} \mid w\} \\ &= 1 - P\{\text{no vehicle at warehouse} \mid w\} \\ &= 1 - (1 - P\{\text{the vehicle at warehouse} \mid w\})^{N_V} \end{aligned}$$

Based on the definition of p_w in Equation (4.12), $p'_{(w,k,0)}$ can be calculated as follows.

$$p'_{(w,k,0)} = 1 - (1 - p_w)^{N_V}, \text{ for } \forall w, k \qquad (4.21)$$

Since the differential equations are changed, the elements in the intensity matrix $Q' = \{q'_{u,v}\}$ can be modified as follows.

$$q'_{(w,k,0),(w+1,k,0)} = \lambda, \; w = 1, 2, \cdots; k = 1, 2, \cdots, w$$

$$q'_{(w,k,0),(0,k,0)} = \lambda_L \left[1 - \left(1 - p_w\right)^{N_V}\right], \; w = 1, 2, \cdots; k = 1, 2, \cdots, w$$

By resolving Equation (3.31) with Q' in above values, the solution for $\pi'_{(0,0,0),t}$ can be obtained. The vehicle utilization is obtained by Equation (3.17), and the transportation cost is the sum of the cost of each vehicle, which is shown in Equation (4.22). The service level S_T can be calculated from Equation (3.28).

$$C_T = \varphi_1 U N_V = \varphi_1 (1 - \pi_{(0,0,0)}) N_V \qquad (4.22)$$

4.3.2. Model Validations and Result Discussions

The following experiments are intended to verify the accuracy of the proposed Markov model for multiple vehicles by comparing results with simulations. In the experiments, the vehicle needs to serve customers who are uniformly distributed in a 100x100 square region. Customers appear in a Poisson manner with an arrival rate $\lambda = 1/16$. Three groups of simulations and Markov model calculations with the number of vehicles N_v equal to 4, 5, and 6 are implemented, respectively. The vehicles which are idle at the warehouse located in the middle of the region have an equal probability to start the delivery trip, once there is pending demand. The loading time at the warehouse is an exponentially distributed random variable with a rate $\lambda_L = 1$. Table 4.3 shows the average results from the simulations, results from the proposed Markov model, and the differences between them.

Table 4.3 compares the simulation and the proposed Markov model in terms of vehicle utilization, average customer waiting time, and service levels. When the number of vehicle running in the service region increases, the vehicle utilization will decrease, since more vehicles share the services for same amount of customers, and each vehicle has less workload as a result. It should be noted that the vehicle utilization in the case of six vehicles is more than 2/3 times of that of four vehicles, due to more efficient routing for longer trips in the case of four vehicles. Meanwhile, the average customer waiting time decreases, while service levels increase. Customers tend to wait for a shorter period, since alternative vehicles can deliver products to new customers when the other vehicles are being utilized. Results in Table 4.3 indicates that when there are more vehicles servicing the same region, the customer waiting time is significantly reduced, however, the costs of transportation and vehicle rental are higher. Service providers need to consider the trade-offs between the quality of services and the cost in order to decide the optimal number of vehicles operating in any given region.

Table 4.3.
Result verification with customer demand rate 1/16

No. of Vehicle		Utilization	Ave Waiting	Service Level			CPU time
				100	300	500	
Simulation	4	0.9603	173.8992	0.3939	0.8416	0.9602	14min
Model		0.9669	159.4443	0.4151	0.8732	0.9743	31.73s
Difference		0.6881	-8.3122	5.3858	3.7460	1.4687	
Simulation	5	0.8589	81.8140	0.7497	0.9787	0.9986	14min
Model		0.8855	70.5757	0.8100	0.9888	0.9996	39.57s
Difference		3.0978	-13.7364	8.0373	1.0327	0.1022	
Simulation	6	0.7541	56.5732	0.8903	0.9953	0.9998	15min
Model		0.7889	45.5324	0.9586	0.9994	0.9999	51.76s
Difference		4.6079	-19.5159	7.6679	0.4174	0.0139	

The symbols and abbreviations are the same as Table 4.1.

The results in the case of five vehicles in Table 4.3 can be compared with the results in the case of only one vehicle with demand rate 1/80 in Table 3.1. The average customer waiting time in the case of five vehicles is nearly half of that in the later case, and the service levels in the case of five vehicles are much higher. The reason is that a new customer tends to wait for shorter time until a vehicle is available to start a delivery trip for him/her in the case of five vehicles.

The proposed model takes much less CPT time to achieve good estimation of various performance measures compared to simulations. When the number of vehicle increases, the CPU time for simulations is relatively constant, while the CPU time for the proposed model increases significantly. It may be due to the efficiency of the heuristic proposed in section 4.3.1 to find out a suitable p_w. In the case with large number of vehicles, a more efficient algorithm should be applied to speed up the calculation of the proposed model.

The differences between results from the Markov model and the simulation are relatively acceptable in terms of vehicle utilization, since they are less than 5%. However, the differences in terms of average customer waiting time and service levels are significant. It can be noted that the difference significantly increases when the number of vehicles increases. This may be due to the assumption that the probability of the other vehicle at the warehouse depends on the number of customers in the waiting list only, but is independent of the status of the observed vehicle. This assumption ignores the relationship between vehicles. Furthermore, the system does not track the observed vehicle any more in the customer waiting process, and the status of that vehicle is eliminated from the system, in order to reduce the state space. The initial transient state probabilities are approximated by the sum of steady state probabilities for different statuses of that vehicle, which is shown in Equation (4.18). The observed vehicle would suddenly become available in the transient state process due to the sum of the probabilities in Equation (4.18). This contradicts with the fact that it is impossible for the observed vehicle to start a new trip immediately when it is in transit. The customer waiting process provides a chance for the vehicle to finish services earlier, and the customer waiting time is estimated to be slightly shorter than the actual waiting time.

In summary, the Markov model is able to approximate the performance of delivery services with multiple vehicles. Service providers are then able to evaluate their business and decide the optimal number of vehicles in a region by using the proposed model. However, the accuracy of the proposed model needs to be improved. More reasonable assumptions and modifications of the model can be further researched.

4.4. Dynamic Pickup and Dynamic Delivery Services

In the real world, goods may need to be transported from customers back to the warehouse. Dynamic pickup services refer to this kind of transportation. Examples of such services include garbage collection, mail/package collection and reverse logistics services. Travelling sales representatives and travelling technician services can also be categorized into dynamic pickup services, since they have similar characteristics.

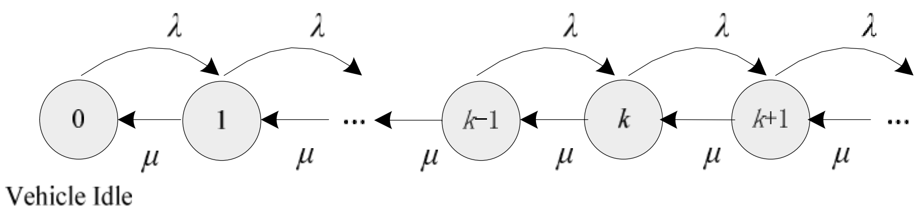
In a static situation, there is no difference between pickup and delivery services, since all information is known in advance. A reverse trip of delivery services is the trip of pickup services. However, if transportation services are considered in dynamic circumstances, there are a few points to consider.

- Firstly, it is unnecessary to take the warehouse into consideration for pickup services if the vehicle has infinite capacity. Once all pending pickups are completed, the vehicle will become idle at the last customer's location. When a new customer appears, the vehicle will move from the current location to the new customer's location. Therefore, there are vehicle routes made from one customer to another customer, but there are no routes between the warehouse and customers.
- Secondly, the routing schedule is more dynamic in pickup services. In delivery services, the routing schedule is fixed after the vehicle starts a trip, and the next routing schedule will be generated only after the vehicle returns. However, in dynamic pickup services, the only fixed schedule is that of the customer site to which the vehicle is currently headed. All subsequent schedules may vary based on the results generated by the routing algorithms.

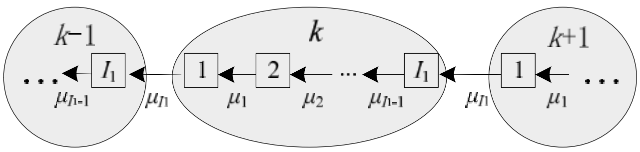
Due to the differences between dynamic pickup and delivery services, the Markov model for dynamic delivery services needs to be modified to adapt the special features of dynamic pickup services.

4.4.1. Markov Model for the Dynamic Pickup Problem

The concept of a waiting list can be ignored in dynamic pickup services, since there are no trips between the warehouse and customers. New demands will be added to the route immediately based on the assumption of the FCFS queue discipline. Customers will be visited one after another according to the scheduled sequence as they appear. The system states is defined as (*k, I*), where *k* is the total number of customers in the routing plan, and *I* is the index of sub-states, which has the same meaning as previously mentioned. Once a new customer's demand appears, *k* will increase by one. When the vehicle travels towards a customer, *I* decreases until zero, which indicates that the vehicle has reached the customer location. State (0,0) indicates the vehicle has become idle since there is no pending customer demand to be fulfilled. The process is illustrated in Fig. 4.8. Fig. 4.8 (b) elaborates the transition from *k* to *k*-1 after several sub steps.



a. Queuing process for dynamic pickup problems



b. Approximate vehicle-travelling process by hypo-exponentially distributed process

Fig. 4.8. Markov Model for the dynamic pickup problem

$\pi = \{\pi_{(k,I)}\}$ denotes the steady state probability of the system. Fig. 4.9 shows the transition diagram based on a certain system state (k, I).

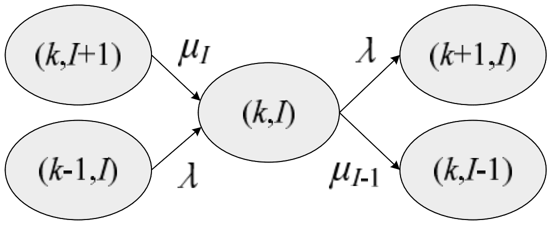


Fig. 4.9. Transition diagram for the dynamic pickup problem

The balance equation for this transition diagram is shown in Equation (4.23).

$$\mu_I \pi_{(k,I+1)} + \lambda \pi_{(k-1,I)} = (\lambda + \mu_{I-1})\pi_{(k,I)} \tag{4.23}$$

The intensity matrix is $Q = \{q_{u,v}\}$. The elements of the intensity matrix are elaborated as follows.

$$q_{(k,I),(k,I-1)} = \mu_{I-1},\ k = 0,1,\cdots; I = 1,2,\cdots, I_1$$

$$q_{(k,0),(k,I_1)} = \mu_{I_1},\ k = 0,1,\cdots$$

$$q_{(k,I),(k+1,I)} = \lambda,\ k = 0,1,\cdots; I = 0,1,\cdots, I_1$$

The diagonal elements of the intensity matrix are calculated as follows.

$$q_{(k,I),(k,I)} = -\sum_{v \neq (k,I)} q_{(k,I),v},\quad \forall (k,I)$$

The steady state probability π is obtained by the stationary equation together with the boundary condition, which are illustrated in Equation (3.16). The vehicle utilization U and transportation cost per unit time C_T are obtained from Equations (3.17) and (3.18).

The customer waiting process starts from a setting time 0, when the specific customer appears. Due to the FCFS service principle, the demand after this specific customer will not affect the waiting time of the specific customer in the system. It is not necessary to consider new demands, and the customer waiting process is a pure death process. The system state is denoted as (k, I) in this process, where k is the position of the specific customer in vehicle route, and I is the index of sub-states. Let $\pi'_t = \{\pi'_{u,t}\}$ be the transient state probability at time t. Initially, the specific customer appears and is scheduled at the end of the vehicle route. The initial state probability $\pi'_{(k,I),0}$ is determined by the steady state probability π as seen in Equation (4.24).

$$\pi'_{(k,I),0} = \pi_{(k-1,I)},\ k = 1,2,\cdots; I = 0,1,\cdots I_1 \tag{4.24}$$

The vehicle visits customers one after another until the specific customer is served. The process ends at state (0, 0), when the demand of the specific customer has been fulfilled. Fig. 4.10 illustrates transition diagram of the process.

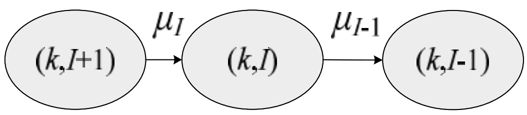


Fig. 4.10. Transition diagram of the transient process

Based on the transition diagram of Fig. 4.10, the transient state equation is formulated as Equation (4.25)

$$\frac{d\pi'_{(k,I),t}}{dt} = \mu_I \pi'_{(k,I+1),t} - \mu_{I-1}\pi'_{(k,I),t} \tag{4.25}$$

The elements of the intensity matrix $Q' = \{q'_{u,v}\}$ are summarized as follows.

$$q'_{(k,I),(k,I-1)} = \mu_{I-1},\ k = 0,1,\cdots; I = 1,2,\cdots,I_1$$

$$q'_{(k,0),(k,I_1)} = \mu_{I_1},\ k = 0,1,\cdots$$

The diagonal elements of the intensity matrix are calculated as follows.

$$q'_{(k,I),(k,I)} = -\sum_{v\neq(k,I)} q'_{(k,I),v},\ \forall(k,I)$$

The solutions for $\pi'_{(0,0),t}$ can be obtained from Equation (3.31). The service level S_T can be calculated by Equation (3.28).

4.4.2. Model Validations and Result Discussions

The following experiments are carried out to verify the accuracy of the proposed Markov model for dynamic pickup services. In the experiments, the vehicle has to fulfill demands uniformly distributed within a 100x100 square region. Customers appear in a Poisson manner. Four groups of simulations and Markov model calculations with customer arrival rate $\lambda = 1/60, 1/80, 1/100, 1/120$ are investigated, respectively. The vehicle is idle at the location of the last served customer, if there is no pending demand. Once a new demand appears, the vehicle travels with a constant speed towards it. The loading and unloading time at customers' locations are not taken into consideration in both simulations and the Markov model. Table 4.4 shows the average results from simulations, results from the proposed Markov model, and the difference between results from these two.

Table 4.4.
Result verification in a 100x100 square region

Demand Rate		Utilization	No. Customer	Service Level		
				100	300	600
Simulation	1/60	0.8792	4.7918	0.2403	0.6380	0.8875
Model		0.8642	4.3213	0.2581	0.6987	0.9337
Difference		-1.7061	-9.8189	7.4074	9.5141	5.2056
Simulation	1/80	0.6579	1.4589	0.5571	0.9515	0.9983
Model		0.6500	1.3938	0.5470	0.9584	0.9989
Difference		-1.2008	-4.4623	-1.8130	0.7252	0.0601
Simulation	1/100	0.5263	0.8939	0.6888	0.9880	0.9998
Model		0.5200	0.8672	0.6724	0.9894	1.0000
Difference		-1.1970	-2.9869	-2.3810	0.1417	0.0200
Simulation	1/120	0.4393	0.6554	0.7581	0.9955	0.9999
Model		0.4333	0.6376	0.7394	0.9961	1.0000
Difference		-1.3658	-2.7159	-2.4667	0.0602	0.0100

"Utilization" represents the usage factor of the vehicle; "No. Customer" represents the average number of customers in the system; "Simulation" represents the results from simulation; "Model" represents the results from calculations using the Markov model; "Difference" represents the difference between the results using the proposed Markov model and the simulation, calculated by 100*(Model-Simulation)/Simulation.

Table 4.4 compares the simulation and the proposed Markov model in terms of vehicle utilization, average number of customers in the system and service levels. Three criterions, 100, 300 and 600 units of time are artificially set to calculate the service levels. In the Markov model, the first two measurements, vehicle utilization and average number of customer, are obtained from the steady state process, while the service levels are obtained from the customer waiting process.

It can be noted that in Table 4.4, when the customer arrival rate increases, the usage of the vehicle will increase, since more customers need service in the same period of time. Meanwhile, the number of customers waiting will sharply increase, and service levels will decrease as well.

Furthermore, Table 4.4 shows that the results obtained from the Markov model are close to those from simulations with most of the differences less than 5%. Even in the heavy traffic case with a 1/60 customer arrival rate, the differences are less than 10%. The errors may result from the non-sufficient state space included in the Markov model. In order to reduce the calculations, the queue length is set to 15, which means a maximum of 15 customers are allowed to wait for services and any subsequent customers will leave since 15 persons are already in the queue. Therefore, the average number of customers obtained from the Markov model is lower than that from the simulation. The service levels from the Markov model are higher than those from the simulation, since the specific customer, which is being tracked to calculate the customer waiting time, never waits until more than 15 persons are served. Extending the queue length will improve the results from the Markov model. When the queue length of the Markov model increases to 25, the difference of average number of customers decreases to 7.69% in the case of

1/60 customer arrival rate, and the differences of the service levels decrease to less than 7%.

The differences between dynamic delivery problems and dynamic pickup problems can be observed when comparing results in Table 4.4 with Table 3.1, which shows the results of dynamic parcel delivery services with similar parameters. The petrol cost in dynamic pickup services is much lower and the service levels are much higher, since the vehicle does not visit the warehouse constantly. It can also be noted that the difference between results from simulations and the proposed model is slightly larger in Table 4.4, which may indicate that the model of dynamic pickup services has less accuracy compared to the model of dynamic delivery services.

4.5. Issue on Routing Strategies

4.5.1. Introduction of Routing Strategies

As previously mentioned, planning parcel transportation services followed the FCFS routing method. This assumption will be discarded in this section. In order to improve the vehicle routing efficiency and reduce transportation costs, service providers may apply some efficient optimization routing strategies in order to generate a better service schedule. Exact algorithms and heuristics are two categories of routing strategies usually used in research.

Heuristics can generate fast and satisfactory solutions. The insertion-based heuristics are usually used to solve DVRP in the literature. A best-insertion algorithm is introduced as follows.

Step 1. Initially, denote $S = \{v_0\}$ as the vehicle route, and $S' = \{v_1, \cdots, v_n\}$ as customers to be inserted in the route.

Step 2. Choose one customer v_i from the set of S', which is nearest to one of customers or the warehouse in the set of S.

Step 3. Try to insert v_i into the vehicle route S with the minimum increase of the total distance of the route. Delete v_i from the set of S'.

Step 4. Repeat Step 2 to 3 until S' becomes empty.

Exact algorithms are usually based on a binary tree search or an integer linear programming. The Branch-and-Bound based algorithms are famous exact algorithms based on the binary tree search. A Branch-and-Bound algorithm with the lower bounds from assignment problems (Laporte, 1992) is described in this section. If there are n nodes representing customer locations and one node representing the warehouse, the size of the distance matrix, which contains the distances between nodes, is $(n+1)\times(n+1)$. The same matrix can be used as a cost matrix in a task assignment problem. The solution of the corresponding assignment problem may not be a suitable solution for the VRP, since it may include sub-trips, which do not include the warehouse. However, the solution generates a good lower bound for the VRP. The following is a Branch-and-Bound algorithm with lower bound generated by the assignment problem.

Step 1. The initial upper bound is provided by the solution from the best-insertion algorithm mentioned previously. The initial lower bound is provided by the famous Hungarian algorithm (Kuhn, 1955) for the corresponding assignment problem with the distance matrix of $n+1$ nodes.

Step 2. Check whether the solution obtained from the assignment problem contains sub-trips. If so, create two branches on the binary tree for a path between two nodes in a sub-trip, one branch with the path forced into the solution and the other with it forced out. The lower bound of each branch is equal to the cost provided by the corresponding assignment problem. If not, a new solution for the VRP is obtained. Each time the Branch-and-Bound algorithm finds a better solution, the solution will become the new upper bound.

Step 3. A branched path, for example (v_i, v_{i+1}), is denoted as 1 or 0. 1 means the path is included in the final solution. The distances from v_i to the other nodes except v_{i+1} and the distances from the other nodes except v_i to v_{i+1} in the distance matrix are set to infinity. 0 means the path in excluded in the final solution. The distances from v_i to v_{i+1} in the distance matrix are set to infinite.

Step 4. The Hungarian Algorithm is applied to generate the solution for the assignment problem with the modified distance matrix in Step 3.

Step 5. Repeat Step 2 to 4. If the lower bound of the branch is higher than or equal to the upper bound, this branch is forfeited. If the lower bounds of all branches are higher than or equal to the upper bound, the optimal solution is achieved.

Details of Branch-and-Bound algorithms and their applications in solving the VRP can be found in the literature (Laporte, 1992; Fisher, 1994). It can be noted that the Branch-and-Bound algorithm schedules all customers equally, and generates a solution regardless of the initial sequence of customers.

These two routing algorithms will be applied in this section to generate vehicle schedules for dynamic parcel delivery services. The following illustrates the Markov model for parcel delivery services with various routing strategies.

4.5.2. Estimation of Vehicle Travel Time

The vehicle travel time between customers and customers, or the warehouse and customers, is definitely different among various routing strategies. τ_0, ${\sigma_0}^2$, τ_1 and ${\sigma_1}^2$ need to be estimated in a similar fashion as in section 3.2, but the value of parameters c_1, c_2, c_3 and c_4 cannot be obtained from Larson and Odoni's research (1981) or calculated by Equation (2.1). The parameters should be estimated by experiments.

In this experiment, randomly generated customers are spread out within a 100×100 square region, where the warehouse is located at the center. Vehicle trips including 1 to 15 customers, respectively, are generated based on two different routing algorithms mentioned previously. The data of vehicle travel time between customers as well as the warehouse is collected for each case with a different number of customers in the trips, respectively. Means and variances of the vehicle travel time are calculated based on this data. Table 4.5 and 4.6 elaborate the expected values and standard deviations of the vehicle travel time.

Table 4.5 shows the means and standard deviations of vehicle travel time based on the Branch-and-Bound routing algorithm, and Table 4.6 shows those values based on the best insertion routing algorithm. The means $\tau_{T,N}$ and standard deviations $\sigma_{T,N}$ of the entire vehicle trip are applied in a steady state process, and $\tau_{0,N}$, $\sigma_{0,N}$, $\tau_{1,N}$ and $\sigma_{1,N}$ are applied in the customer waiting process in the Markov model. It can be noted that the mean of the overall travel time of the vehicle trip is $\tau_{T,N} = 2\tau_{0,N} + (N-1)\tau_{1,N}$, since a vehicle trip is composed of several paths between customers and two paths between the warehouse and a customer. However, the variance of the vehicle trip $\sigma_{T,N}^2 \neq 2\sigma_{0,N}^2 + (N-1)\sigma_{1,N}^2$, since the travel time of two successive paths are correlated with each other in the case of certain routing strategies.

Table 4.5.
Travel time between nodes (Branch-and-Bound Algorithm)

N	$\tau_{1,N}$	$\sigma_{1,N}$	$\tau_{0,N}$	$\sigma_{0,N}$	$\tau_{T,N}$	$\sigma_{T,N}$
1			38.2729	14.2306	76.5458	28.4614
2	52.0374	24.7892	38.2387	14.2667	128.5149	38.6993
3	45.7662	23.1382	37.1109	14.6167	165.7542	41.1468
4	41.0340	21.7570	35.6046	15.0392	194.3110	41.4461
5	37.4035	20.5044	34.0502	15.3452	217.7143	40.8627
6	34.5460	19.3456	32.1827	15.5112	237.0951	39.8304
7	32.1959	18.3658	30.6250	15.5817	254.4255	38.8220
8	30.2423	17.5150	28.9669	15.4643	269.6299	37.8719
9	28.5832	16.7350	27.3582	15.2124	283.3822	37.0878
10	27.1465	16.0267	26.0397	14.9415	296.3981	36.2448
11	25.8700	15.3869	24.7159	14.6027	308.1315	35.3973
12	24.7440	14.8051	23.5462	14.2137	319.2766	34.6986
13	23.8023	14.2747	22.3683	13.6875	330.3646	33.9965
14	22.8565	13.7973	21.4252	13.1585	339.9855	33.5508
15	22.0742	13.3674	20.3956	12.6809	349.8305	32.8792

"N" represents the number of customers visited in a vehicle trip. "$\tau_{1,N}$" is the mean value of travel time between any two customers in a vehicle trip with N customers, and "$\sigma_{1,N}$" is the standard deviation. "$\tau_{0,N}$" is the mean value of travel time between the warehouse and a customer in a vehicle trip, and "$\sigma_{0,N}$" is the standard deviation. "$\tau_{T,N}$" represents the mean value of travel time of a entire vehicle trip, and "$\sigma_{T,N}$" represents the standard deviation.

In section 3.2, the vehicle travel time was approximated by the hypo-exponential distribution, which was composed of a series of Poisson distributed travel time. However, in Table 4.5 and 4.6, the means and variances of the vehicle travel time between customers are different among the vehicle trips with different number of customers included. This was not observed in sections 3.2 and 3.3. The previous model is not suitable, and the system state cannot be denoted as (w,k,I) any more. One more dimension of the state h is needed, which will indicate the total number of customers included on the trip. For each h, a specific hypo-exponential distribution needs to be applied to approximate the vehicle travel time. However, the four dimensions in the system state make the state space too large and overload the computation effort. The structure of the Markov model needs to be modified to reduce the state space.

Table 4.6.
Travel time between nodes (Best-Insertion Algorithm)

N	$\tau_{1,N}$	$\sigma_{1,N}$	$\tau_{0,N}$	$\sigma_{0,N}$	$\tau_{T,N}$	$\sigma_{T,N}$
1			38.3139	14.2029	76.6279	28.4062
2	52.2840	24.8181	38.2202	14.2659	128.7244	38.8034
3	45.9503	23.3982	37.1846	14.5471	166.2697	41.6396
4	40.7196	21.3548	36.5168	15.1092	195.1925	42.1416
5	37.2945	19.9799	35.9634	15.4959	221.1047	42.1639
6	34.5131	18.6072	34.9374	15.8471	242.4404	41.4148
7	32.3021	17.6426	34.0425	16.1142	261.8976	41.2869
8	30.5317	16.7984	32.9378	16.2224	279.5971	40.3026
9	29.0151	16.0455	32.1238	16.3488	296.3683	40.0731
10	27.7748	15.5176	31.0464	16.3684	312.0656	39.5143
11	26.6806	14.9985	29.9033	16.1896	326.6127	39.0215
12	25.7389	14.5499	28.9608	16.0305	341.0490	39.3075
13	24.8358	14.1678	27.8961	15.7716	353.8212	39.1564
14	24.0844	13.8273	27.0177	15.6562	367.1326	38.8015
15	23.3948	13.5168	26.1311	15.3101	379.7890	38.5652

The symbols and abbreviations are the same as Table 4.5.

Firstly, a hyper-hypo-exponential distribution is introduced as follows. The hyper-hypo-exponential distribution is obtained by arranging several different hypo-exponential distributions in parallel. Fig. 4.11 shows a general example of a hyper-hypo-exponential distribution. In Fig. 4.11, n parallel levels are depicted and each level j contains a series of r_j phases of exponentially distributed travel time. Each level j is selected with probability p_j, and the summation of p_j is one.

Fig. 4.12 (a) is another example, which shows a special structure of the hyper-hypo-exponential distribution. In this special case, there are j phases of exponentially distributed travel time in level j, and the j-1 phases are the same as those in level j-1. The structure can be further simplified as Fig. 4.12 (b). The structure of hyper-hypo-exponentially distributed travel time in Fig. 4.12 (b) is the basis for the Markov model to approximate real vehicle travel time in this section.

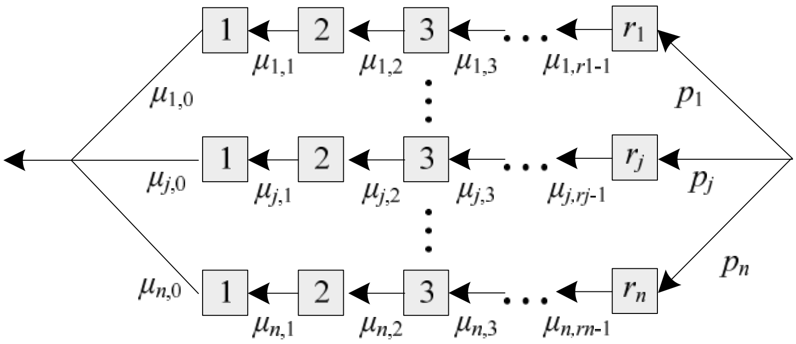
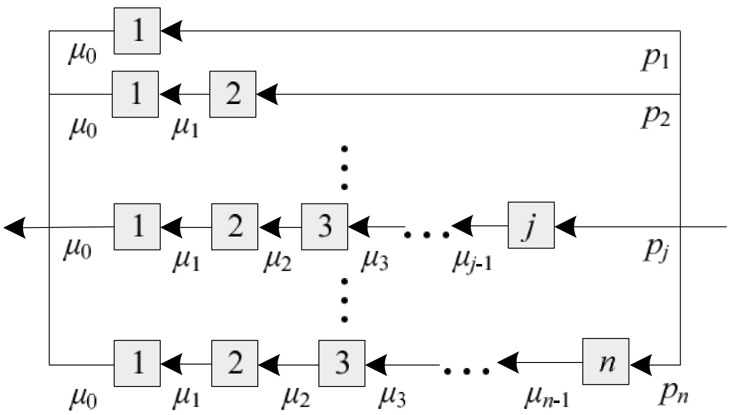
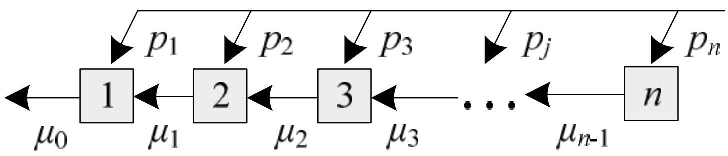


Fig. 4.11. An example of hyper-hypo-exponentially distributed travel time



a. An special case of hyper-hypo-exponentially distributed travel time



b. The equivalent structure of (a)

Fig. 4.12. The hyper-hypo-exponentially distributed travel time in the Markov model

The aim of this exercise is to construct a universal process with one set of $\{\mu_j\}$ as shown in Fig. 4.12 (b), modelling the vehicle en route, and to generate different hyper-hypo-exponentially distributed travel time with different sets of $\{p_j\}$ in order to approximate the exact vehicle travel time of trips with different number of customers. The system state in the steady state process can be denoted as (w,I), where w represents the number of customers in the waiting list and I is the indicator of sub-states of a vehicle trip. The trip is no longer divided into several paths between customers in the steady state process.

The mean $\tau_{T,N}$ and variance $\sigma_{T,N}^2$ of the hyper-hypo-exponentially distributed travel time in Fig. 4.12 (b) are formulated as Equation (4.26) and Equation (4.27), respectively.

$$\tau_{T,N} = \sum_{j=1}^{n} p_{N,j} \sum_{k=0}^{j-1} \frac{1}{\mu_k} \tag{4.26}$$

$$\sigma_{T,N}^2 = \sum_{j=1}^{n} p_{N,j} \sum_{k=0}^{j-1} \frac{1}{\mu_k^2} + \sum_{j=1}^{n} p_{N,j} \left(\sum_{k=0}^{j-1} \frac{1}{\mu_k} \right)^2 - \left(\sum_{j=1}^{n} p_{N,j} \sum_{k=0}^{j-1} \frac{1}{\mu_k} \right)^2 \tag{4.27}$$

$p_{N,j}$ is the probability that level j is selected in the hyper-hypo-exponentially distributed travel time for the trip with N customers.

Given a certain value of $\tau_{T,N}$ and a set of $\{\mu_j\}$, the maximum value of variance is obtained from the solution of Equation (4.28) with $p_{N,j} = 0$ for all $2 \le j \le n-1$. (Fig. 4.13)

$$\tau_{T,N} = p_{N,1} \frac{1}{\mu_0} + p_{N,n} \sum_{k=0}^{n-1} \frac{1}{\mu_k} \tag{4.28}$$

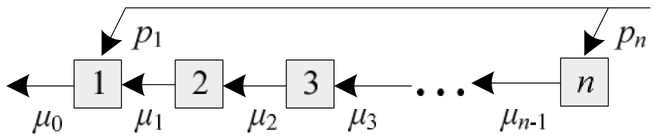


Fig. 4.13. Vehicle travel time with maximum variance

The maximum value of variance is

$$\sigma_{T,N}^2 = \frac{2}{\mu_1^2} \frac{\sum_{k=0}^{n-1} \frac{1}{\mu_k} - \tau_{T,N}}{\sum_{k=0}^{n-1} \frac{1}{\mu_k} - \frac{1}{\mu_1}} + \frac{\tau_{T,N} - \frac{1}{\mu_1}}{\sum_{k=0}^{n-1} \frac{1}{\mu_k} - \frac{1}{\mu_1}} \sum_{k=0}^{n-1} \frac{1}{\mu_k^2} + \frac{\tau_{T,N} - \frac{1}{\mu_1}}{\sum_{k=0}^{n-1} \frac{1}{\mu_k} - \frac{1}{\mu_1}} \left(\sum_{k=0}^{n-1} \frac{1}{\mu_k} \right)^2 - \tau_{T,N}^2$$

The maximum variance is obtained from the summation of four elements. The first three elements are increasing functions of $\frac{1}{\mu_{n-1}}$, respectively. If the mean value is fixed, the maximum variance is an increasing function of $\frac{1}{\mu_{n-1}}$.

Similarly, the minimum variance is obtained from the solution of Equation (4.29) with $p_{N,j} = 0$ for all $j \le i-2$ or $j \ge i+1$, where i is obtained from $\sum_{k=0}^{i-2} \frac{1}{\mu_k} < \tau_{T,N} \le \sum_{k=0}^{i-1} \frac{1}{\mu_k}$. (Fig. 4.14.)

$$\tau_{T,N} = p_{N,i-1} \sum_{k=0}^{i-2} \frac{1}{\mu_k} + p_{N,i} \sum_{k=0}^{i-1} \frac{1}{\mu_k} \tag{4.29}$$

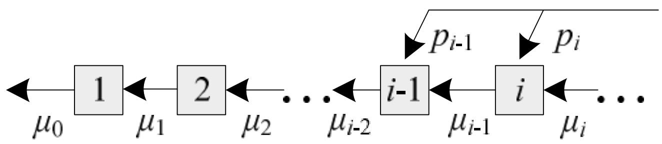


Fig. 4.14. Vehicle travel time with minimum variance

Proposition 4.1: Given fixed values of $\tau_{T,N}$, part of the set ($\mu_0, \mu_1, \cdots, \mu_{i-2}$), and $\tau_{T,N} > \sum_{k=0}^{i-2} \frac{1}{\mu_k}$. If two more elements: μ_{i-1} and μ_i, have to be added to satisfy $\sum_{k=0}^{i-1} \frac{1}{\mu_k} < \tau_{T,N} \le \sum_{k=0}^{i} \frac{1}{\mu_k}$, the variance of travel time is able to achieve minimum value when

$$\frac{1}{\mu_i} = \frac{1}{\mu_{i-1}} = \frac{1}{2}\left(\tau_{T,N} - \sum_{k=0}^{i-2} \frac{1}{\mu_k} \right) \tag{4.30}$$

Proof:

The minimum variance with fixed $\tau_{T,N}$ and ($\mu_0, \mu_1, \cdots, \mu_{i-2}, \mu_{i-1}, \mu_i$) is obtained based on Equation (4.29).

$$\tau_{T,N} = (1 - p_{N,i+1}) \left(\sum_{k}^{i-2} \frac{1}{\mu_k} + \frac{1}{\mu_{i-1}} \right) + p_{N,i+1} \left(\sum_{k}^{i-2} \frac{1}{\mu_k} + \frac{1}{\mu_{i-1}} + \frac{1}{\mu_i} \right)$$

The solution for $p_{N,i+1}$ is as follows.

$$p_{N,i+1} = \left(\tau_{T,N} - \sum_{k}^{i-2} \frac{1}{\mu_k} - \frac{1}{\mu_{i-1}} \right) \Big/ \frac{1}{\mu_i} \qquad (4.31)$$

$p_{N,i+1}$ is substituted by $p_{N,i} = 1 - p_{N,i+1}$, and $p_{N,j} = 0$ for all $j \neq i, i+1$ into Equation (4.27).

The corresponding variance is calculated as follows.

$$\sigma_{T,N}^2 = \sum_{k=0}^{i} \frac{1}{\mu_k^2} - \left(\tau_{T,N} - \sum_{k=0}^{i} \frac{1}{\mu_k} \right)^2 \qquad (4.32)$$

Equation (4.31) can be rewritten as follows.

$$\sigma_{T,N}^2 = \sum_{k=0}^{i-1} \frac{1}{\mu_k^2} - \left(\tau_{T,N} - \sum_{k=0}^{i-1} \frac{1}{\mu_k} \right)^2 + 2\frac{1}{\mu_i} \left(\tau_{T,N} - \sum_{k=0}^{i-1} \frac{1}{\mu_k} \right)$$

Since $\sum_{k=0}^{i-1} \frac{1}{\mu_k} < \tau_{T,N} \leq \sum_{k=0}^{i} \frac{1}{\mu_k}$, $\sigma_{T,N}^2$ is an increasing function of $\frac{1}{\mu_i}$, if $\tau_{T,N}$ and ($\mu_0, \mu_1, \cdots, \mu_{i-2}$, μ_{i-1}) are fixed. Therefore, $\sigma_{T,N}^2$ achieves its minimum value when $\frac{1}{\mu_i} = \tau_{T,N} - \sum_{k=0}^{i-1} \frac{1}{\mu_k}$. This value of $\frac{1}{\mu_i}$ is substituted into Equation (4.32).

$$\sigma_{T,N}^2 = \sum_{k=0}^{i-2} \frac{1}{\mu_k^2} + \frac{1}{2}\left(\tau_{T,N} - \sum_{k=0}^{i-2} \frac{1}{\mu_k} \right)^2 + 2\left(\frac{1}{\mu_{i-1}} - \frac{1}{2}\tau_{T,N} + \frac{1}{2}\sum_{k=0}^{i-2} \frac{1}{\mu_k} \right)^2$$

Therefore, $\sigma_{T,N}^2$ achieves its minimum value when the following equation holds.

$$\frac{1}{\mu_i} = \frac{1}{\mu_{i-1}} = \frac{1}{2}\left(\tau_{T,N} - \sum_{k=0}^{i-2} \frac{1}{\mu_k} \right)$$

The proposition 4.1 is proved and the minimum $\sigma_{T,N}^2$ is as follows.

$$\sigma_{T,N}^2 = \sum_{k=0}^{i-2} \frac{1}{\mu_k^2} + \frac{1}{2}\left(\tau_{T,N} - \sum_{k=0}^{i-2} \frac{1}{\mu_k} \right)^2$$

The following is an algorithm to find a set of $\{\mu_j\}$ suitable for all vehicle trips, as well as a number of sets of $\{p_{N,j}\}$.

Step 1. Initialization. Follow the algorithm in section 3.2 to find a solution of Equation (3.5) for $\tau_{T,1}$ and $\sigma_{T,1}^2$, denoted as $\{\mu_0, \mu_1, \cdots, \mu_{i-2}\}$. Let $N = 1$.

Step 2. Let $N = N + 1$. Find the minimum variance for trip travel time with *N* customers where the mean is equal to $\tau_{T,N}$. Add two more elements, μ_{i-1} and μ_i, into the set of $\{\mu_j\}$. The values of μ_{i-1} and μ_i are obtained from Equation (4.30).

Step 3. Check whether the minimum variance is suitable. If $\sum_{k=0}^{i} \frac{1}{\mu_k^2} > \sigma_{T,N}^2$, find the maximum element in the set of $\{\mu_j\}$, and break it into two elements with equal values. Repeat Step 3 until $\sum_{k=0}^{i} \frac{1}{\mu_k^2} \le \sigma_{T,N}^2$.

Step 4. Repeat Step 2 and 3, until $N = 15$, if the queue length is set to 15.

Step 5. Add an element μ_n with a large value to the end of the set of $\{\mu_j\}$, and make sure that the maximum variance for the trip with *N* customers is greater than $\sigma_{T,N}^2$ for all *N* from 1 to 15, when the mean is equal to $\tau_{T,N}$.

Step 6. The set of $\{\mu_j\}$ is suitable for all vehicle trips, where *N* increases from 1 to 15, since the minimum variances which can be generated from the set of $\{\mu_j\}$ are less than $\sigma_{T,N}^2$, and the maximum variances are greater than $\sigma_{T,N}^2$. Suitable sets of $\{p_{N,j}\}$ must be found by solving Equation (4.26) and Equation (4.27) for all *N* from 1 to 15.

4.5.3. Estimation of Transportation Costs in Steady State Process

In this section, transportation costs are investigated in the steady state process. The system state is denoted as (w, I), and the steady state probability is $\pi = \left\{\pi_{(w,I)}\right\}$. One set of $\{\mu_j\}$ and 15 sets of $\{p_{N,j}\}$ are formulated into transition rates in balance equations and the intensity matrix $Q = \left\{q_{u,v}\right\}$.

1) The vehicle returns to the warehouse and starts the next trip (State $(w,1)$)

   When the vehicle finishes its current trip, the vehicle starts its next trip to serve the waiting customer immediately. As in the previous assumption, the vehicle spends no time at the warehouse loading and unloading products and thus the states $(w,0)$ are redundant states when $w \geq 1$. Since a hyper-hypo-exponentially distributed vehicle travel time is used to approximate the time the vehicle spend on a trip, the process transitions from state $(w,1)$ to state $(0,I)$ with probability $\mu_0 p_{w,I}$, where $I = 1,2,\cdots,n$.

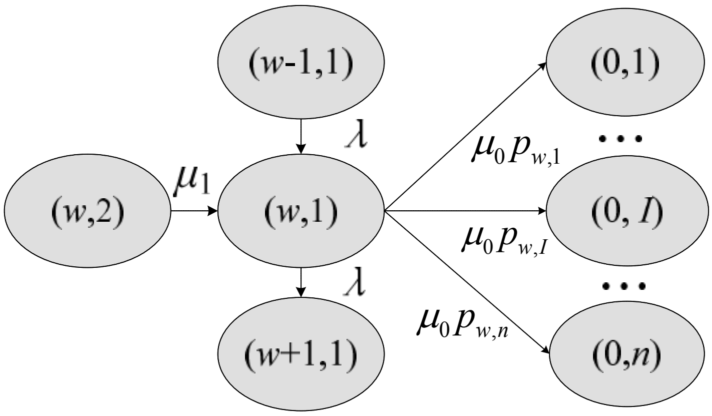


Fig. 4.15. Transition diagram in the situation that the vehicle returns to the warehouse and starts the next trip

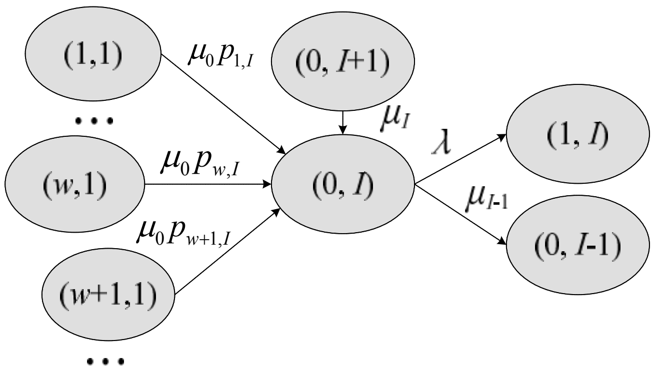
Based on the transition diagram in Fig. 4.15, the balance equation is formulated as Equation (4.33)

$$\mu_1 \pi_{(w,2)} + \lambda \pi_{(w-1,1)} = (\mu_0 + \lambda)\pi_{(w,1)} \tag{4.33}$$

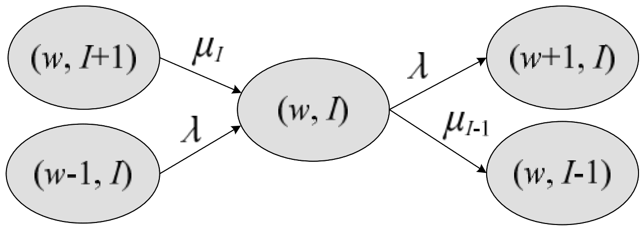
The left side represents the total transition rate from other states to state (w,1), which represents the vehicle travelling and demand appearing. The right side of Equation (4.33) indicates the total transition rate from state (w,1) to other states. The transition out of the state (w,1) happens when the vehicle starts the next trip or a new customer demand appears. It can be noted that $\sum_{I=1}^{n} \mu_0 p_{w,I} = \mu_0$, since $\sum_{I=1}^{n} p_{w,I} = 1$.

2) The vehicle is travelling on the road. (State (w,I))

As previously mentioned, there is a probability $\mu_0 p_{w,I}$ that the process transitions from state (w,1) to state $(0, I)$ for all $w > 0$, which is illustrated in Fig 4.16 (a). In the case of more than one customers queued in the waiting list, the transition is similar to that in section 3.3 (Fig 4.16 (b)).



a. Transition diagram based on system state $(0, I)$



b. Transition diagram of vehicle travelling with more than one customer in waiting list

Fig. 4.16. Transition diagram of the vehicle travelling on the road

The balance equations for these two cases are as follows.

$$\sum_{w=1}^{15} \mu_0 p_{w,I} \pi_{(w,1)} + \mu_I \pi_{(0,I+1)} = (\mu_{I-1} + \lambda)\pi_{(0,I)} \tag{4.34}$$

$$\lambda \pi_{(w-1,I)} + \mu_I \pi_{(w,I+1)} = (\mu_{I-1} + \lambda)\pi_{(w,I)} \tag{4.35}$$

3) The vehicle is idle at the warehouse. (State (0,0))

When there are no customers waiting for services, the vehicle is idle at the warehouse. Once new demand appears, the process transitions from state $(0,0)$ to state $(0,I)$ with probability $p_{w,I}$, where $I = 1, 2, \cdots, n$. The transition diagram for this situation is illustrated in Fig. 4.17. The corresponding balance equation is shown as Equation (4.36). It can be noted that $\sum_{I=1}^{n} \lambda p_{1,I} = \lambda$, since $\sum_{I=1}^{n} p_{1,I} = 1$.

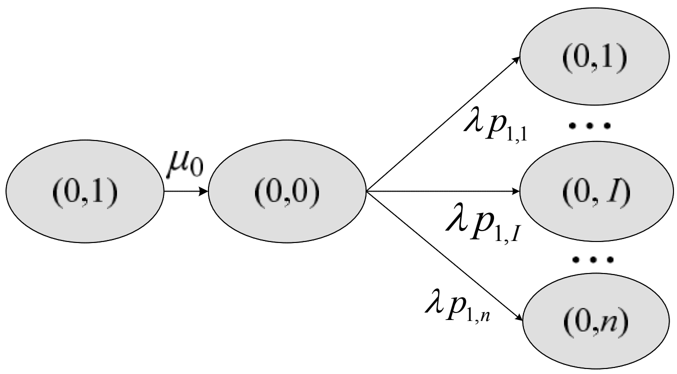


Fig. 4.17. Transition diagram of vehicle idle at the warehouse

$$\mu_0 \pi_{(0,1)} = \lambda \pi_{(0,0)} \tag{4.36}$$

All the balance equations are summarized in the stationary equation $\pi Q = 0$, and the elements in the intensity matrix $Q = \{q_{u,v}\}$ are listed in Appendix A.3.1.

The steady state probability π is obtained by solving the stationary equation together with the boundary condition $\sum \pi = 1$, similar to solutions in section 3.3. The vehicle utilization U is obtained in Equation (4.37).

$$U = \sum_{u \neq (0,0)} \pi_u = 1 - \pi_{(0,0)} \tag{4.37}$$

The transportation cost per unit time C_T is proportional to the vehicle utilization.

$$C_T = \varphi_1 U = \varphi_1 (1 - \pi_{(0,0)}) \tag{4.38}$$

In Equation (4.38), φ_1 is the petrol cost per unit time when vehicle is in transit.

4.5.4. Estimation of Service Levels in Transient Customer Waiting Process

The customer waiting process is applied to estimate the waiting time of a specific customer. This process starts from time 0, when the system is stable and this specific customer appears. It ends when the service for this specific customer finishes and the customer leave the system. The customer waiting process is divided into two steps. The first step is when the vehicle is travelling on the current trip, and the specific customer is waiting for his/her delivery to be scheduled. The transition in this step is similar to the one in steady state process, until the vehicle reaches the warehouse. The other step is the new vehicle trip which includes the specific customer. In the second step, the demand of the specific customer is fulfilled.

Since the service does not follow FCFS queue discipline, the position of the specific customer in the routing schedule is not related to the sequence of the customers. An example based on an insertion algorithm is illustrated as follows. There are already j customers on the next trip when the specific customer appears. Without any specific information, this customer could be allotted in any position in the schedule with equal probability. Therefore, the probability that the specific customer is arranged in i^{th} position is $p_i = \frac{1}{j+1}$. At a later time t, another new customer will be inserted into the schedule. The probability that the new customer is arranged before the specific customer is $\frac{i}{j+2}$. The specific customer is in i^{th} position, only when the specific customer is in i^{th} position before time t and the new customer is inserted after him/her, or the specific customer is in $(i\text{-}1)^{\text{th}}$ position before time t and the new customer is inserted before him/her. Therefore, the probability that the specific customer is finally located in the i^{th} position is $p_i = \frac{1}{j+1}\left(1-\frac{i}{j+2}\right)+\frac{1}{j+1}\frac{i-1}{j+2} = \frac{1}{j+2}$. In summary, it is fair to assume that the specific customer has an equal chance of being slotted in any position of the second trip, as long as the routing strategy does not favor specific customers. This can be contrasted with the case where the FCFS strategy favors customers appearing early.

In the steady state process, the vehicle trip is no longer divided into paths between each two customers, and only the travel time for an entire vehicle trip is considered. However, in the customer waiting process, the position of the specific customer on the second vehicle trip is important. Estimating travel time for the entire trip is not adequate to measure the waiting time of the specific customer. Therefore, an extra dimension has to be added to the system state in order to indicate the position of the specific customer in the second vehicle trip. The system state is defined as (w,k,I), where the first and third symbols have the same meaning with the previous steady state process, and k represents the position of the specific customer on the second vehicle trip. Since k is only valid for the specific customer, it is set to -1 in order to indicate that the vehicle is still on the first trip. In the first trip, the vehicle follows the schedule and any new demands will be cumulated in the waiting list. After the vehicle reaches the warehouse, it starts the second trip with the specific customer included in the schedule. This specific customer has an equal chance to be served in any order.

In the second step, the schedule for the specific customer is determined. New demand will not affect the waiting time of the specific customer and will not be taken into consideration. w becomes the indicator of the vehicle trip with w customers. The structure of the model for the second trip is similar to that in section 3.4. However, the travel time for the trip with varying number of customers is different. $\tau_{1,N}$ and $\sigma_{1,N}$ are used to denote the mean and standard deviation of the travel time between two customers on the trip with N customers, which are listed in Table 4.5 and 4.6. $\mu_{N,i}$ denotes the transition rate from state (N,k,i+1) to (N,k,i), where $k \geq 1$. $\tau_{0,N}$ and $\sigma_{0,N}$ denote the mean and standard deviation of the travel time between the warehouse and a customer on the trip including N customers. $\mu'_{N,i}$ denotes the transition rate from state (N,0,i+1) to (N,0,i). Vehicle travel time is approximated by the hypo-exponentially distributed random variables, which satisfy the following equations. A feasible solution of $\mu_{N,i}$ and $\mu'_{N,i}$ can be obtained from the algorithm in section 3.2.

$$\begin{cases} \tau_{1,N} = \sum_{i=0}^{I_{1,N}-1} \frac{1}{\mu_{N,i}} \\ \sigma_{1,N}^2 = \sum_{i=0}^{I_{1,N}-1} \frac{1}{\mu_{N,i}^2} \\ \tau_{0,N} = \sum_{i=0}^{I_{0,N}-1} \frac{1}{\mu'_{N,i}} \\ \sigma_{0,N}^2 = \sum_{i=0}^{I_{0,N}-1} \frac{1}{\mu'^2_{N,i}} \end{cases}$$

Let $\pi'_t = \{\pi'_{u,t}\}$ be the transient state probability at time t, where $\pi'_{u,t}$ is the transient state probability of state u at time t, and $Q' = \{q'_{u,v}\}$ the intensity matrix, where $q'_{u,v}$ is the transition rate from state u to v in the customer waiting process.

Initially, at time 0, the specific customer appears and joins others in the waiting list. The initial state probability $\pi'_{(w,-1,I),0}$ is determined by the steady state probability π obtained from previous process, due to the assumption that the system is stable at time 0.

$$\pi'_{(w,-1,I),0} = \pi_{(w-1,I)}, \quad w = 1,2,\cdots; I = 1,2,\cdots,n \tag{4.39}$$

If the vehicle is initially idle at the warehouse, the specific customer will immediately be served.

$$\pi'_{(1,1,0),0} = \pi_{(0,0)} \tag{4.40}$$

The following illustrates transition diagrams and corresponding differential equations for every situation.

1) The vehicle is travelling on the road of the first trip. (State (w,-1,I))

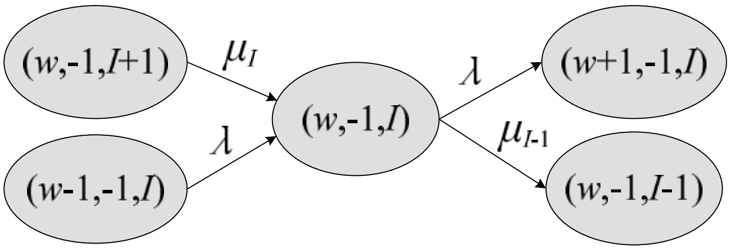


Fig. 4.18. Transition diagram of vehicle travelling on the road of the first trip

$$\frac{d\pi'_{(w,-1,I),t}}{dt} = \mu_I \pi'_{(w,-1,I+1),t} + \lambda \pi'_{(w-1,-1,I),t} - (\lambda + \mu_{I-1})\pi'_{(w,-1,I),t} \tag{4.41}$$

2) The vehicle returns to the warehouse and starts the second trip. (State (w,-1,1))

As previously mentioned, the state (w,-1,0) is redundant, since the vehicle starts the second trip once it returns to the warehouse. The system transitions from state (w,-1,1) to (w,k,0) with equal chances of $1/w$.

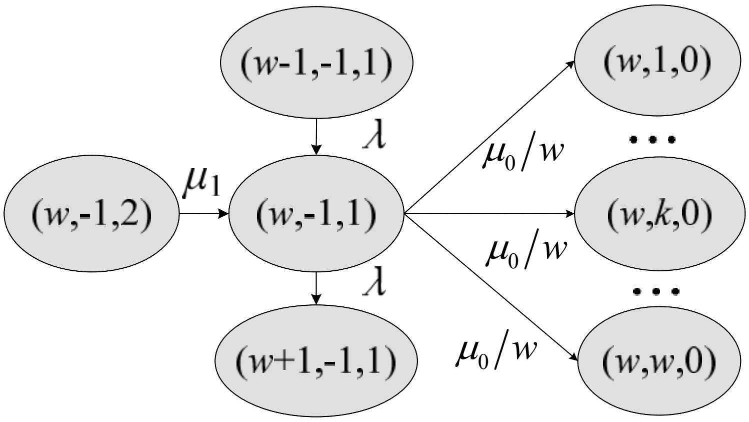


Fig. 4.19. Transition diagram of vehicle returning to the warehouse

$$\frac{d\pi'_{(w,-1,1),t}}{dt} = \mu_1\pi'_{(w,-1,2),t} + \lambda\pi'_{(w-1,-1,1),t} - (\lambda + \mu_0)\pi'_{(w,-1,1),t} \tag{4.42}$$

3) The vehicle is travelling between customers on the second trip. (State (w,k,I))

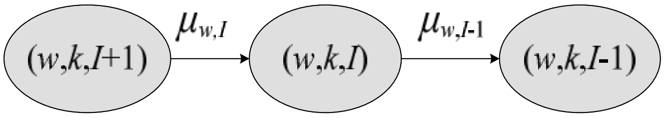


Fig. 4.20. Transition diagram of vehicle travelling between customers on the second trip

$$\frac{d\pi'_{(w,k,I),t}}{dt} = \mu_{w,I}\pi'_{(w,k,I+1),t} - \mu_{w,I-1}\pi'_{(w,k,I),t} \tag{4.43}$$

4) The vehicle reaching and leaving a customer on the second trip. (State (w,k,0))

The specific customer may have a chance to be served in the k^{th} position of the second trip. The transition rate from state (w,-1,1) to (w,k,0) is μ_0/w.

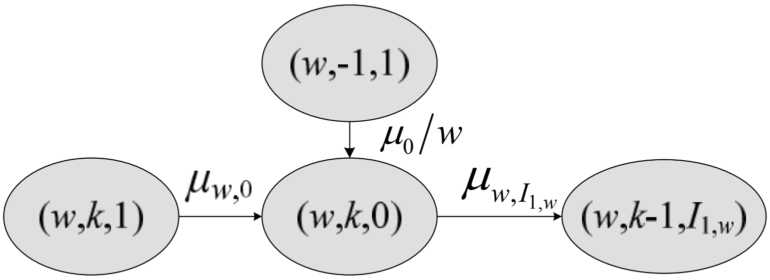


Fig. 4.21. Transition diagram of vehicle leaving a customer on the second trip

$$\frac{d\pi'_{(w,k,0),t}}{dt} = \mu_{w,0}\pi'_{(w,k,1),t} + \frac{\mu_0}{w}\pi'_{(w,-1,1),t} - \mu_{w,I_{1,w}}\pi'_{(w,k,0),t} \quad (4.44)$$

5) The vehicle is travelling to the warehouse on the second trip. (State (w,0,I))

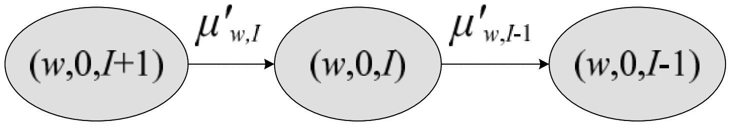


Fig. 4.22. Transition diagram of vehicle heading to the warehouse on the second trip

The transition rate from state (w,0,I) to state (w,0,I-1) becomes $\mu'_{w,I-1}$, when the trip includes w customers.

$$\frac{d\pi'_{(w,0,I),t}}{dt} = \mu'_{w,I}\pi'_{(w,0,I+1),t} - \mu'_{w,I-1}\pi'_{(w,0,I),t} \quad (4.45)$$

6) The vehicle finishes the service for the specific customer. (State (0,0,0))

Once the vehicle trip for the specific customer is completed, the process terminates at state (0,0,0), which is illustrated in Fig. 4.23.

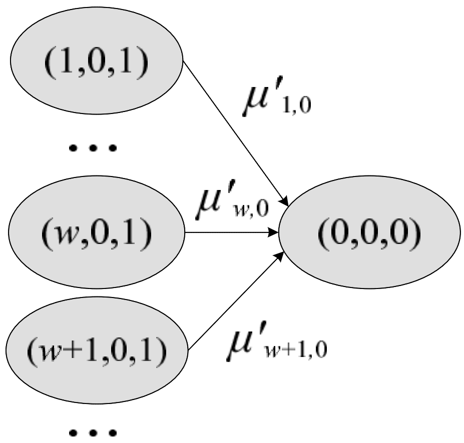


Fig. 4.23. Transition diagram of the end of the process

$$\frac{d\pi'_{(0,0,0),t}}{dt} = \sum_{w} \mu'_{w,0}\pi'_{(w,0,1),t} \quad (4.46)$$

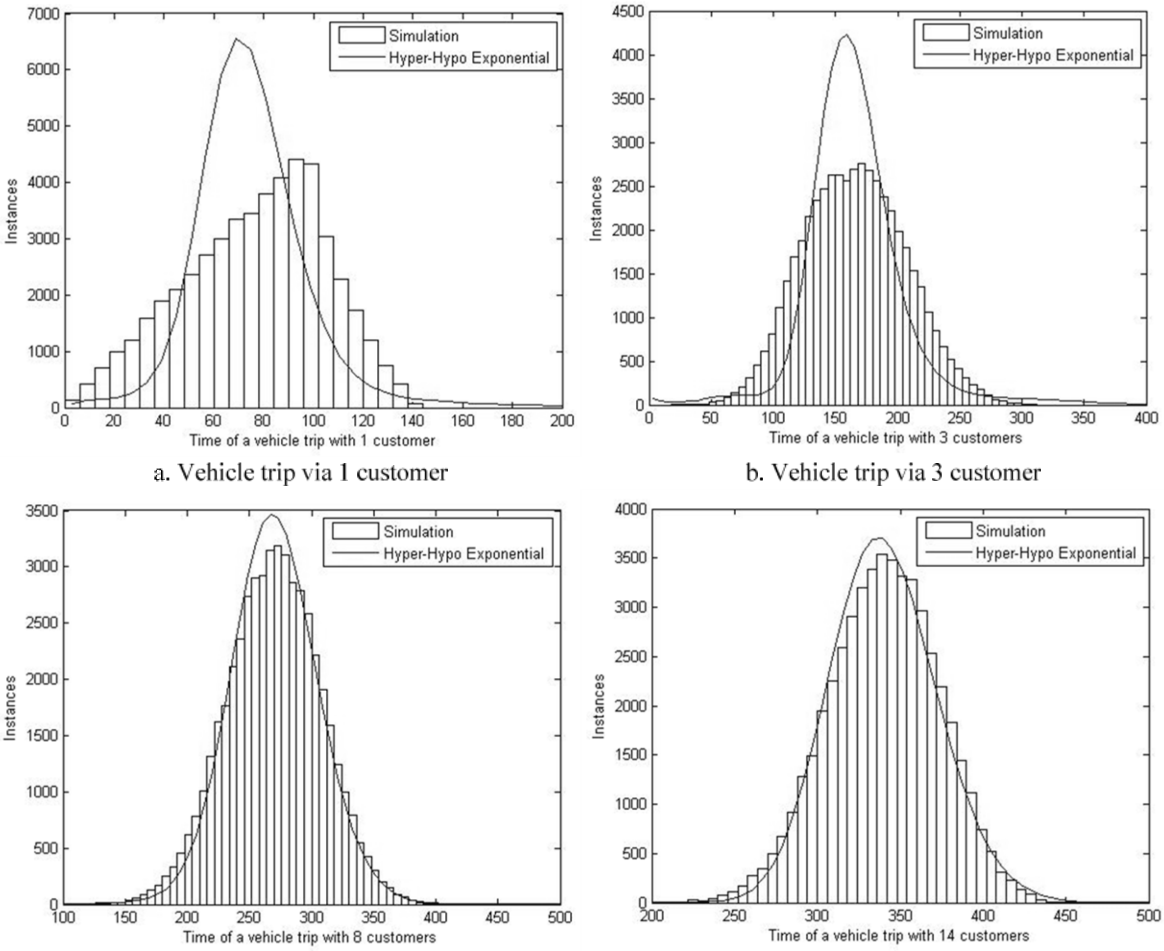
The solutions of the above differential equations can be obtained by Equation (3.31). The elements in the intensity matrix are summarized in Appendix A.3.2.

The solution of $\pi'_{(0,0,0),t}$ can be obtained at any time t based on Equation (3.31), subject to the initial conditions (4.39) and (4.40). $\pi'_{(0,0,0),t}$ is the cumulative distribution function of the waiting time of the specific customer. Finally, the service level S_T is obtained from Equation (3.28).

4.5.5. Model Validations

Firstly, the approximation of the hyper-hypo-exponentially distributed vehicle travel time is evaluated. Fig. 4.24. shows the histograms and corresponding density curves of hyper-hypo-exponentially distributed vehicle travel time for trips including 1, 3, 8 and 14 customers, respectively. The vehicle trips are generated by the Branch-and-Bound algorithm mentioned previously. The data of the histograms is collected from simulations. For example, in the case of eight customers, the simulation generates 50,000 trips. Each trip starts from the warehouse, passes eight randomly generated customers according to the sequence suggested by the Branch-and-Bound algorithm, and ends at the warehouse. The vehicle travel time of trips is segregated into small intervals, and the numbers of instances in the intervals are illustrated in Fig. 4.24 (c) as histograms. The mean and variance of the travel time from these 50,000 trips are applied to generate the density curves of hyper-hypo-exponential distribution, based on the algorithms mentioned in section 4.4.3. This curve is also illustrated in Fig. 4.24 (c).

In Fig. 4.24, the vehicle travel time of the trip with one or three customers is not fit for the hyper-hypo-exponential distribution. When there are more customers included in a trip, the distribution of the travel time becomes close to the hyper-hypo distribution, but the approximation of the travel time is still not accurate. The errors of the approximations would contribute to the final errors in estimating the transportation costs and customer waiting time.



a. Vehicle trip via 1 customer

b. Vehicle trip via 3 customer

c. Vehicle trip via 8 customer

d. Vehicle trip via 14 customer

Fig. 4.24. The distribution of the vehicle travel time of a trip via several customers (Branch-and-Bound Algorithm)

The following experiments are intended to verify the accuracy of the proposed Markov model by comparing results with simulations. The simulations are similar to those in section 3.6, but the routing strategies are no longer FCFS. Two routing algorithms, the Branch-and-Bound algorithm and the best-insertion algorithm are investigated. Solving the Markov model is more efficient since it only takes 40 seconds, while the simulation takes more than an hour.

The Branch-and-Bound algorithm is investigated in the first set of experiments. The vehicle needs to serve customers who are uniformly distributed in a 100x100 square region. Customers appear in a Poisson manner with an arrival rate λ. Five groups of simulations and Markov model calculations with $\lambda = 1/30, 1/40, 1/50, 1/60, 1/80$ are implemented, respectively. The vehicle starts the delivery trip from the warehouse located in the middle of the region, travels with a constant speed on the delivery route, and returns to the warehouse after the current trip to load products for next delivery trip. The routing plans are generated by the Branch-and-Bound algorithm. The loading and unloading time at the warehouse or customers' locations are not taken into consideration. Table 4.7 shows the average results from simulations, results from the proposed Markov model, and the difference between them.

Table 4.7.
Result validation for Branch-and-Bound Algorithm

Demand Rate		Utilization	Ave. Waiting	Service Level			CPU time
				100	300	500	
Simulation	1/30	0.9998	296.4064	0.0643	0.5279	0.9264	19min
Model		0.9997	292.7695	0.0617	0.5412	0.9338	19.85s
Difference		-0.0168	-1.2270	-4.0750	2.5217	0.7925	
Simulation	1/40	0.9928	234.3832	0.1265	0.7223	0.9830	18min
Model		0.9940	236.6592	0.1154	0.722	0.9802	19.55s
Difference		0.1202	0.9711	-8.7597	-0.0412	-0.2781	
Simulation	1/50	0.9684	190.9717	0.2164	0.8409	0.9949	20min
Model		0.9697	191.8245	0.2082	0.8457	0.9936	20.54s
Difference		0.1359	0.4465	-3.7737	0.5688	-0.1284	
Simulation	1/60	0.9237	158.7747	0.3142	0.9099	0.9987	21min
Model		0.9233	158.3083	0.3180	0.9142	0.9973	20.08s
Difference		-0.0452	-0.2938	1.1845	0.4680	-0.1358	
Simulation	1/80	0.7982	115.6197	0.5034	0.9712	0.9999	22min
Model		0.8016	116.6580	0.5079	0.9688	0.9992	19.24s
Difference		0.4197	0.8980	0.8944	-0.2413	-0.0696	

The symbols and abbreviations are the same as Table 4.1.

Table 4.7 compares the simulation and the proposed Markov model in terms of vehicle utilization, average customer waiting time and service levels. In the Markov model, the first measurement, utilization, is obtained from the steady state process mentioned in section 4.4.3, while the average customer waiting time and service levels are obtained from the customer waiting process mentioned in section 4.4.4.

In terms of CPU time, the proposed model is much faster than simulations to make estimation of various performance measures. When the demand rates decreases, the CPU time of the model is relatively constant, while the simulation time slightly increases, since system tends to generate vehicle trips visiting fewer customers and more departure and arrival events at the depot in the case of low demand rate. The differences between results from the Markov model and the simulation are relatively small in terms of vehicle utilization. They are less than 0.5%, since the model for steady state process is accurate. Fig. 4.25 shows the steady state probability in the case of customer demand 1/40. It illustrates the probability that there are 0~15 new customers in line for parcel delivery services in the waiting list. The curves of probability from the Markov model and the simulation almost overlap with each other, which indicates that the Markov model can approach the vehicle utilization and transportation costs accurately in the steady state process.

Estimating performance measures from the customer waiting process is also accurate in the case of λ=1/40. Fig. 4.26 illustrates the cumulative distribution of the customer waiting time. Although, the difference between the Markov model and the simulation are visible at the beginning of the curves, the two curves are still consistent with each other. The transient state Markov process is able to accurately estimate the customer waiting time. Table 4.7 shows that the average customer waiting time and service levels obtained from the Markov model are close to those from simulations. Most of the differences are less than 4%, and the maximum difference is 8.76%. Large differences occur when the required service time is less than 100 units of time, which is difficult to achieve by service providers.

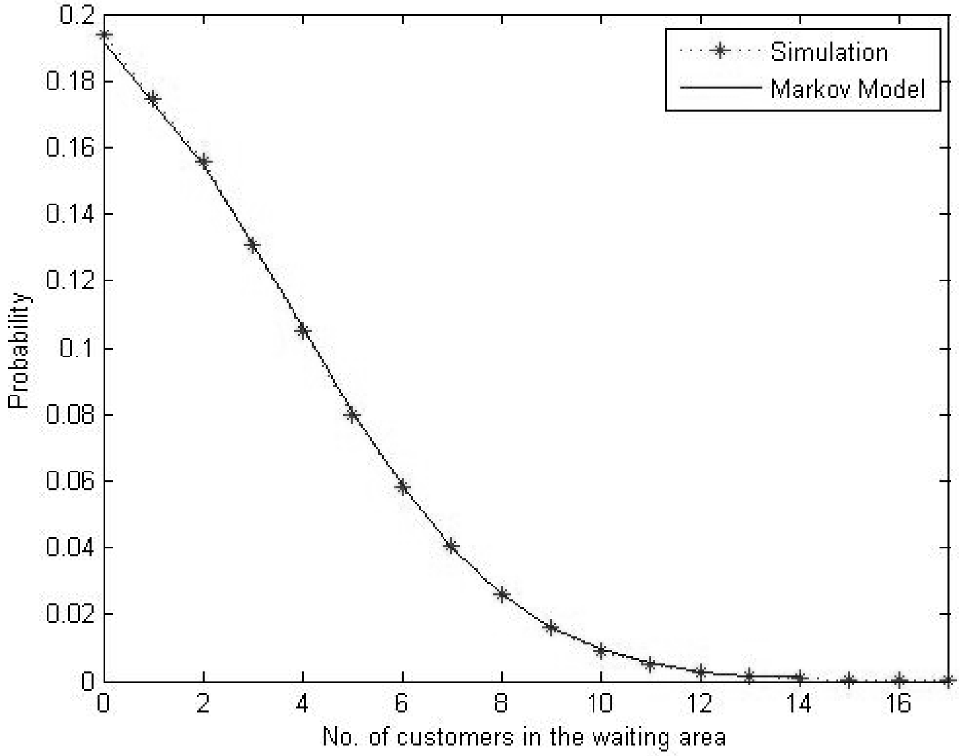


Fig. 4.25. Steady State probability of customers in waiting list (λ=1/40)

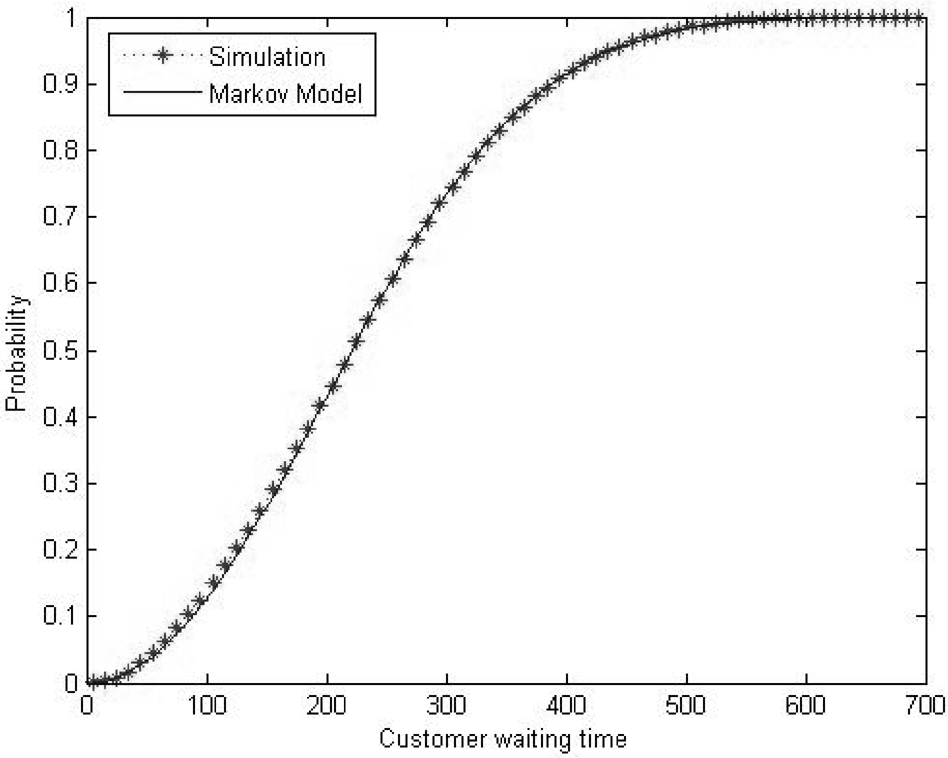


Fig. 4.26. Cumulative distribution of Customer waiting time (λ=1/40)

Errors may result from the approximation of the vehicle travel time. As previously mentioned, the vehicle travel time is not accurately fit for the hyper-hypo-exponential distribution. Especially in the cases where there are a small amount of customers on a vehicle trip, the errors are distinctly large. That is the reason why the final errors are relatively large in the cases with small demand rates, since the system in this case tends to generate vehicle trips with a small amount of customers. However, it is unexpected that the errors in the final results in terms of transportation costs and customer waiting time are relatively small. It indicates that the results generated by the Markov model do not significantly rely on the accuracy of the estimation of the vehicle travel time. A good estimation of the transportation costs and customer waiting time can be achieved by providing the mean and variance of the travel time.

The errors may also come from the non-sufficient state space involving in the Markov model in large demand cases. In order to reduce computation effort, the queue length has been set to 15. It is more likely that the number of customers in the waiting list may exceed the queue length in the cases with large demand rates. Therefore, the average number of customers obtained from the Markov model may be lower and the service levels may be higher than those from the simulation. Extending the queue length may improve the result from the Markov model. However, extending the queue length will significantly increase the state space, since one of the dimensions I in the system state must satisfy $I \geq \frac{\tau_{T,N}^2}{\sigma_{T,N}^2}$, and $\lim_{N\to\infty} \frac{\tau_{T,N}^2}{\sigma_{T,N}^2} = \frac{\beta A}{0} \to \infty$ (see, Beardwood et al., 1959; Steele, 1981; and Lawler et al., 1985). This indicates that there are limitations of the Markov model when the demand is large.

In the second set of experiments, every condition is the same as the previous experiments except that the schedules of the vehicle trips are generated by a best-insertion algorithm. The results are shown in Table 4.8, which are similar to those in Table 4.7. However, the differences between the results from the proposed Markov model and the simulation are relatively larger. The Markov model estimates the performance measure better for the Branch-and-Bound algorithm than the best-insertion algorithm.

In terms of CPU time, the conclusion is similar to that in the experiments of Branch-and-bound algorithms. The proposed model is much faster than simulations. When the demand rates decreases, the CPU time of the model is relatively constant, while the simulation time slightly increases. Compared to the Branch-and-bound algorithms, the simulation time is shorter, since the best-insertion algorithm generates vehicle routes faster. In the calculation of the proposed model, the CPU time for the insertion algorithm is larger than that of Branch-and-bound algorithms, since the customer waiting process of the insertion algorithm has larger number of states based on the parameters in tables 4.5 and 4.6.

Table 4.8.
Result validation for Best-Insertion Algorithm

Demand Rate		Utilization	Ave. Waiting	Service Level			CPU time
				100	300	600	
Simulation	1/30	0.9998	335.6659	0.0401	0.4299	0.8600	15min
Model		0.9998	319.6652	0.0489	0.4670	0.8881	22.82s
Difference		-0.0053	-4.7669	21.8492	8.6363	3.2638	
Simulation	1/40	0.9944	256.4294	0.0902	0.6619	0.9652	17min
Model		0.9947	250.5525	0.1023	0.6783	0.9671	22.66s
Difference		0.0282	-2.2918	13.5001	2.4824	0.1938	
Simulation	1/50	0.9699	203.8252	0.1763	0.8139	0.9912	19min
Model		0.9712	198.2840	0.1979	0.8263	0.9903	22.57s
Difference		0.1250	-2.7186	12.2644	1.5180	-0.0884	
Simulation	1/60	0.9245	166.6762	0.2851	0.8967	0.9974	20min
Model		0.9249	161.4838	0.3106	0.9058	0.9963	22.31s
Difference		0.0426	-3.1153	8.9463	1.0091	-0.1135	
Simulation	1/80	0.7956	118.9122	0.4823	0.9689	0.9999	22min
Model		0.8029	117.6698	0.5040	0.9669	0.9989	22.49s
Difference		0.9153	-1.0448	4.4878	-0.2148	-0.1031	

The symbols and abbreviations are the same as Table 4.1.

Table 4.8 shows that the results obtained from the Markov model are close to those from simulations. Especially for vehicle utilization, the differences are within 1%. Most of the differences in results generated from customer waiting process are less than 5%. Large errors happen when the required service time is less than 100 units of time, which is difficult to achieve by service providers. It is obvious that the demand in the case of demand rate 1/30 may exceed the limit of the queue length, since the vehicle utilization is close to 100% and customers are waiting for a long time. That is why the errors are large in this case. Extending the queue length may improve the results.

Comparing Table 4.7 and 4.8 with Table 3.1, the routing strategies based on the Branch-and-Bound algorithm and the best-insertion algorithm generate more efficient vehicle routes for parcel delivery services. The transportation costs are thus reduced. Especially, the customer waiting time has been significantly reduced and the service levels have been improved as well. Furthermore, results generates from the Branch-and-Bound algorithm are slightly better than those from the best-insertion algorithm.

In summary, results generated from the Markov model are close to those from simulation. The Markov model provides results much faster than simulations. The computation time are spent on finding solutions for Equations (3.16) and (3.31). A deliberate design and improvement of the intensity matrices would further increase the efficiency of the proposed model. Therefore, the Markov model can replace the simulations to quickly and accurately estimate performance measures for parcel transportation services with various routing strategies.

5. Model Applications in Management Decisions

5.1. Overview

In chapter 3 and 4, the proposed Markov models were established to estimate performance measures for parcel transportation services. Based on the estimations, service providers are able to systematically evaluate their business, and make management decisions in order to improve performance. In this chapter, three management decision-making examples are discussed. In section 5.2, service price is determined in order to maximize profit on delivery services. In section 5.3, the network design for parcel transportation services is analyzed. The size of the service region and the minimum number of warehouses as well as their locations are decided using optimization algorithms. In section 5.4, an order acceptance problem is analyzed. In this kind of problem, an estimated delivery time is provided to potential customers before they make orders. That delivery time will affect the customers' decisions. A suitable approach to decide the estimated delivery time is discussed in the section.

5.2. Pricing Problems for Parcel Delivery Services

5.2.1. Description of Pricing Problems

In the past ten years, a number of supermarkets and fast food retailers, including Amazon.com, Carrefour and McDonalds, have launched online ordering systems. This has allowed customers to order products from "virtual" storefronts via the Internet or over the telephone and get them delivered with additional charges. In some cases, the retailers' logistics departments provide delivery services, but most of these services are outsourced to third party logistics companies. Outsourcing the logistics function allows retailers to handle complicated logistics issues in a more effective manner. However, this also creates a "service pricing problem" in which these outsourcing companies or internal logistics departments need to choose prices carefully and reduce transportation costs effectively while maintaining excellent service quality. The goal of this section is to provide an efficient method to help outsourcing companies or internal logistics departments address the service-pricing problem in a dynamic vehicle routing based environment. In this type of environment, a single trip with multiple delivery stops provides economics of scale.

Nowadays, the competition for delivery services is fierce, especially concerning the issue of price (Potvin et al., 2006). By examining pricing strategy carefully, a company is able to adjust its realized demand and operating costs to achieve maximum profit. The logistics company has to carefully decide the price that is acceptable to most customers, yet also make delivery services profitable.

In the literature, optimizing the price for products and services has received wide attention, including published works by Chen et al. (2001), Abad and Jaggi (2003), Shinn and Hwang (2003), and Viswanathan and Wang (2003). Various pricing policies are discussed in a supplier-buyer channel, based on the assumption that demand is a decreasing function of the product price. The relationship between price and demand can be linear (You, 2006; Zhou and Lee, 2009), convex functional, such as $\lambda = \alpha / x^2$ (Lei et al. 2006), or even more general, like $\lambda = \alpha x^{-b}$ (Monahan et al. 2004), where the demand rate λ is a commonly used convex function of the price x and α is a random variable with a known distribution function. Dong et al. (2009), Li et al. (2009) and Asdemir et al. (2009) used the multinomial logit (MNL) model to analyze the relationship between demand and price. The MNL model, which is based on discrete choice theory, provides a reasonable way to describe a customer's choice behavior among several alternatives. The incorporation of the MNL model in a stochastic demand process can well explain how price affects demand.

Papers on pricing problems are seldom based on transportation services. Zhou and Lee (2009) analyzed the problem of empty equipment repositioning between two fixed places. Li and Liu (2008) agreed with Hill et al. (2002) that other than the service price, the service quality is another important factor that affects customer decisions. Li and Liu assumed that the demand rate presents a linear increase along with a drop of price and response time, which includes manufacturing and transportation time. All these papers tried to solve the differential equations of demand and find the best price such that demand or revenue is maximized. However, for logistics companies, setting the price in order to maximize sales or revenue only, and ignoring operating costs and service level, can result in great losses. Operating costs and service levels, two significant factors contributing to profit, should be taken into consideration when managers make decisions.

Geunes et al. (2007) published the first paper about the pricing problem in a dynamic vehicle routing based environment. They used continuous approximation methods to obtain vehicle travelling costs and constructed the profit function of the delivery service as the difference between revenue and travelling costs. This paper tried to discuss the relationship between the price and the final profit. Without any numerical results, it is difficult to evaluate the accuracy and efficiency of their methods. Furthermore, their method is difficult to extend to deal with customer waiting time and service levels. Effective evaluation models to predict cost and service level are still needed, as they are the challenging parts to resolve in the pricing problem for parcel delivery services.

This research effectively satisfies the requirement of the pricing problem in parcel delivery services. In the proposed method, customer information, parcel delivery service conditions and routing strategies must be provided. A discrete choice model has been chosen to describe the relationship between the demand and the service price. The proposed Markov models are used to estimate the travelling costs and service levels. A simple yet efficient optimization algorithm has been applied to find the optimal service price in order to maximize profit.

5.2.2. Discrete Choice Model

A discrete choice model is used to describe the choices that people make among a finite set of alternatives. It statistically relates the choices of each person to the attributes of that person and the attributes of the alternatives available to that person. Manski and McFandden (1981) developed the discrete choice model and won the Nobel Prize in 2000. The application of this model is elaborated as follows to estimate the actual demands for parcel delivery services. The price of the service, the reputation of the service provider, the income of the customer and other factors will affect a customer's choice. A customer's choice can be measured using a function with all the effects of these factors. Since this research is restricted to the effect of the price for delivery services, it is assumed that the effects other than the price x is a random variable ε that follows a Logistic distribution. Thus, the measure of a customer's choice can be expressed as $-\beta x+\varepsilon$, where $-\beta x$ is the effect of the service price x, β is the coefficient of the price and ε is the effect of other factors. It is assumed that a customer will make an order when the measure is greater than a constant U_0. The probability that a customer will make an order for delivery services is as follows.

$$\text{Prob}\{-\beta x+\varepsilon > U_0\} = \text{Prob}\{\varepsilon > U_0+\beta x\} = \frac{1}{1+e^{U_0+\beta x}} \tag{5.1}$$

With this probability, the relationship between the service price and demand can be developed. It is assumed that customer arrivals follow a Poisson process with an arrival rate Λ. Only a portion of potential customers seeking delivery services may accept the price. Those customers making orders are the demand λ, which is a function of the service price.

$$\lambda = \Lambda \cdot \text{Prob}\{-\beta x+\varepsilon > U_0\} = \frac{\Lambda}{1+e^{U_0+\beta x}} \tag{5.2}$$

5.2.3. Optimization of the Pricing Problem

Profit consists of the difference between revenue and the sum of travelling cost and penalty for overdue service. Based on the results from the previous sections, the revenue, the travelling cost and the penalty can be calculated as Equations (5.3), (5.4) and (5.5), respectively. All of them are functions of the service price x. Therefore, the profit is a function of the price x. They are determined as follows.

$$\text{Revenue} = \lambda x = \frac{\Lambda x}{1+e^{U_0+\beta x}} = f_1(x) \tag{5.3}$$

$$\text{Travelling Cost} = \varphi_1 U = f_2(x) \tag{5.4}$$

$$\text{Penalty} = \varphi_2(1-S_T)\lambda x = f_3(x) \tag{5.5}$$

$$\text{Profit} = \text{Revenue} - \text{Travelling Cost} - \text{Penalty} = f_1(x) - f_2(x) - f_3(x) = f(x)$$

Based on the results from the discrete choice model and the Markov models, the profit function can be formulated as follows.

$$f(x) = \frac{\Lambda x}{1+e^{U_0+\beta x}} - \varphi_1 U - \varphi_2\left(1-S_T\right)\frac{\Lambda x}{1+e^{U_0+\beta x}} \tag{5.6}$$

In Equation (5.6), φ_1 is the petrol cost per unit time when the vehicle travels on the road and φ_2 is the penalty for an overdue service as a percentage of the service price.

A golden section search algorithm is used to find the optimal service price, with which the profit of this service is maximized. The golden section search is a classic and fast one-dimension search algorithm in textbooks (Press et al. 2007), and it still has broad applications in recent research (Zuo et al. 2011; Patron et al. 2012; Du et al. 2012). The algorithm for this pricing problem is illustrated as follows.

Step 1. Choose an initial price. The revenue function is increasing initially and then is decreasing, while the cost and the penalty functions are monotonically decreasing. This means that the optimal price must be greater than the price at which revenue is maximized. Therefore, the initial price is chosen at the point that revenue is maximized, which is a Lambert's W function $x_1 = [W(e^{-1-U_0}) + 1]/\beta$.

Step 2. Choose a one-step search length *L*, and let $x_2 = x_1 + L$, $x_3 = x_1 + L/(1+\xi)$, where ξ is called the golden ratio ($\xi \approx 1.618$). Find the profit for the three prices, $f(x_1)$, $f(x_2)$ and $f(x_3)$. If $f(x_1) > f(x_3) > f(x_2)$, let $L = L/(1+\xi)$. If $f(x_1) < f(x_3) < f(x_2)$, let $L = (1+\xi)L$. Repeat step 2 until $f(x_3) > f(x_1), f(x_2)$.

Step 3. Search within the region $\Re = [x_1, x_1 + L]$. Let $x_3 = x_1 + L/(1+\xi)$ and $x_4 = x_1 + \xi L/(1+\xi)$. Find the profit for the two prices, $f(x_3)$ and $f(x_4)$. If $f(x_3) > f(x_4)$, let $L = \xi L/(1+\xi)$. If $f(x_3) < f(x_4)$, let $L = \xi L/(1+\xi)$ and $x_1 = x_3$. Repeat step 3 until $L < \varepsilon$, where ε is a relatively small real number.

Step 4. Finally, the optimal price is x_3 if $f(x_3)$ is greater, or x_4 if $f(x_4)$ is greater.

It is noted that golden section search is a one-dimension search technique for finding the extremum of a strictly unimodal function, which has only one peak without any other local maxima or minima. Therefore, the application of golden section search algorithm for the pricing problem implies that the profit is a unimodal function of the price, and experiments shows that the objective function has only one peak. It is aim to illustrate the applications of the proposed Markov model in the pricing problems. A better optimization algorithm is open for future research of pricing problems for dynamic parcel delivery services.

5.2.4. Results and Discussions

Experiments are carried out to verify the accuracy of the proposed pricing model for parcel delivery services. The results from the proposed approach are compared with the results from simulations. The performance measures of revenue, vehicle utilization, service level and profit are evaluated. Both simulation and the calculation of the proposed pricing model are implemented in MATLAB 7.0 on a PC with 2.33GHz CPU and 3.25GB of RAM. Solving the Markov model only cost several seconds, while the simulation takes a few minutes. The proposed Markov model is much faster than the simulation. In the first experiment, the vehicle delivers products to customers who are uniformly distributed in a 100x100 square region. Customers appear in a Poisson manner with an inter-arrival rate Λ=1/40. The delivery charge is \$30.00 per customer. The percentage of customers who are willing to pay this amount for the delivery service is calculated based on the discrete choice model. The service provider promises that customer demands will be fulfilled within 300 units of time; otherwise, the customer will acquire compensation equal to 1/3 of the service price for each overdue service. It is assumed that the vehicle starts a delivery trip from the warehouse located in the middle of the region, travels at a constant speed with a petrol cost of $0.30 per unit time and returns to the warehouse after the current trip to load products for the next trip. Table 5.1 shows the results of the vehicle starting a delivery trip with at least one to six pending demands.

Table 5.1 compares the simulation and the Markov model in terms of revenue, vehicle utilization, service level and profit. The simulation results are the average values from ten instances of experiments, and the percentage below the number is the 95% confidence interval for those results. In the proposed approach, the first measurement, the expected revenue is obtained from the Poisson customer arrival rate and the discrete choice model. It is the same value in all six cases, since the customer arrival rate and service price are fixed. However, revenue is varied in simulation due to random errors.

The vehicle utilization is obtained from the steady state process mentioned in section 3.3. It reflects the vehicle travelling cost per unit time. The difference between the results from the Markov model and the simulation is less than 1%, since the Markov model for this process is more accurate.

The service level is obtained from the customer waiting process mentioned in section 3.4. This process is based on the results from the steady state process, and it has more assumptions and approximations. Therefore, the differences for service levels are higher, as much as 6%, but still acceptable. The profit is calculated based on Equation (5.6), which is the combination of the values from the previous three measurements. The differences for profit are at the same level as service levels.

Table 5.1.
Results when the service price is $30.00

	Departure Criterion	Revenue	Utilization	Service Level	Profit	CPU Time
Simulation	1	0.37529 ±0.066%	0.830121 ± 0.050%	0.886234 ± 0.084%	0.112024 ± 0.186%	21min
Model		0.37500	0.830984	0.905082	0.113983	7.60s
Difference		-0.0779%	0.1039%	2.1268%	1.7481%	
Simulation	2	0.37513 ±0.105%	0.772775 ± 0.084%	0.847886 ± 0.127%	0.124281 ± 0.197%	15min
Model		0.37500	0.770034	0.867943	0.127681	10.73s
Difference		-0.0347%	-0.3547%	2.3655%	2.7361%	
Simulation	3	0.374980 ±0.107%	0.743014 ± 0.113%	0.750908 ± 0.216%	0.120941 ± 0.201%	12min
Model		0.37500	0.739250	0.773777	0.125221	12.22s
Difference		0.0512%	-0.5066%	3.0456%	3.5387%	
Simulation	4	0.37499 ±0.063%	0.725813 ± 0.085%	0.586396 ± 0.146%	0.105548 ± 0.142%	11min
Model		0.37500	0.720795	0.612392	0.110539	15.75s
Difference		0.0027%	-0.6914%	4.4332%	4.7284%	
Simulation	5	0.37519 ±0.097%	0.714623 ± 0.113%	0.391133 ± 0.288%	0.084654 ± 0.175%	10min
Model		0.37500	0.708535	0.412592	0.089065	19.13s
Difference		-0.0506%	-0.8519%	5.4864%	5.2103%	
Simulation	6	0.37522 ±0.066%	0.705984 ±0.087%	0.215001 ±0.562%	0.065241 ±0.298%	9min
Model		0.37500	0.699806	0.224179	0.067888	22.85s
Difference		-0.0584%	-0.8750%	4.2689%	4.0563%	

"Departure criterion" represents the minimum number of customers visited in a vehicle trip; "Revenue" represents the income of the service per unit time; "Utilization" represents the vehicle travelling cost; "Service Level" represents the percentage of delivery tasks completed within 300 units of time. "Simulation" represents the results from simulation; "Model" represents the results from calculations using the Markov model; "Difference" represents the difference between the results using the Markov model and the simulation, calculated by 100*(Model-Simulation)/Simulation.

In Table 5.1, as the vehicle departure criterion increases, the vehicle utilization will increase and the service level will decrease, since the vehicle is idle for a longer period at the warehouse, waiting for enough demands to start the next trip. With more customers on a trip, the vehicle is able to save transportation costs. Therefore, the profit may increase initially as the vehicle departure criterion increases. In addition, the confidence intervals become larger as the vehicle departure criterion increases. Using service level as an example, the confidence interval is almost 1% narrower in the case of a departure criterion of one versus a departure criterion of six. The differences between the results from the Markov model and the simulations also become larger, in general. For example, the difference in profit is 3.5% lower in the case of a departure criterion of one versus a departure criterion of five.

In summary, the Markov model generates results much faster than simulations. Therefore, the Markov model is able to quickly and accurately estimate performance measures for parcel transportation services. Based on these estimations, pricing algorithms can be used to decide a suitable price for the service.

In the following experiment, the best price to maximize total profit for parcel delivery services is found. All of the parameters are the same as the previous experiment except that the service price is a random variable. The golden section search algorithm mentioned in section 5.2.3 will be applied to decide the best price for the service. Table 5.2 shows the optimal results for each departure criterion.

The golden section search finds the optimal solution for each departure criterion, and the best price obtained is different in each case. For a departure criterion of one to three, the best price is decreasing, but it is increasing as the departure criterion increases from three.

From Table 5.2, the maximum profit is achieved when the vehicle is allowed to start a new delivery trip with at least two pending demands, and the service price is set at \$34.79. As long as the departure criterion increases, the vehicle utilization decreases, which indicates transportation costs are decreasing, and the service level decreases, which indicates penalties for overdue services are increasing. The overdue penalty is relatively small at the beginning and has minimal effect on profit. Transportation costs are reduced as the departure criterion increases, and total profit increases accordingly. If vehicle departure criterion is more than three, the service level dramatically decreases. The overdue penalty is the main factor that reduces total profit.

Table 5.2.
Optimal results for the pricing problem

Departure Criterion	Price	Revenue	Utilization	Service Level	Profit	CPU Time
1	34.95	0.330872	0.672785	0.983101	0.127240	183.16s
2	34.79	0.332690	0.605224	0.935651	0.144130	219.44s
3	34.38	0.337091	0.587543	0.818649	0.140690	285.41s
4	34.43	0.336538	0.568048	0.634815	0.125379	369.05s
5	35.16	0.328599	0.532748	0.417083	0.105052	468.06s
6	36.38	0.314437	0.485935	0.214158	0.086200	545.28s

The symbols and abbreviations are the same as Table 5.1.

The computational time to obtain a satisfactory solution is less than ten minutes, which is acceptable. The CPU time increases as the departure criterion increases, since more effort is needed to estimate performance measures, which is shown in Table 5.1.

The maximum profit obtained from the golden section search algorithm is illustrated in Fig. 5.1. They are compared with the results in the case of a fixed service price of \$25.57. The \$25.57 price is obtained from maximizing revenue only, since most of the pricing researches have focused on maximizing sales and revenue.

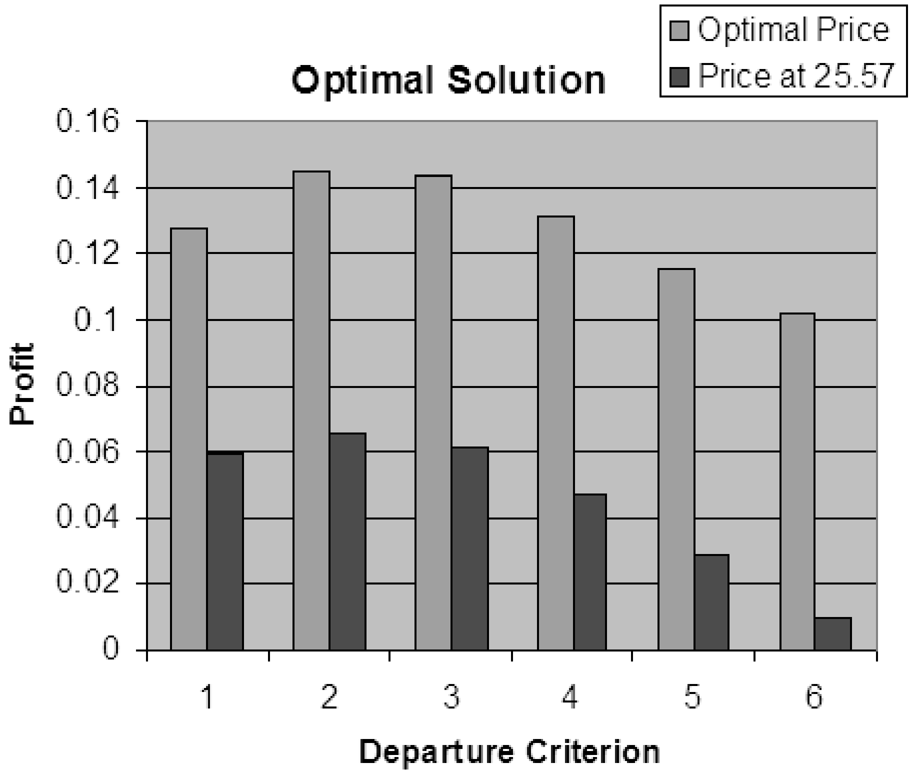


Fig. 5.1. Profit comparison between the optimal price and a fixed price of $25.57

It is obvious that applying the best price obtained from the proposed pricing process will significantly improve the profit for parcel delivery services. For example, the optimal profit in the case with a departure criterion of one is more than $0.12. Compared to $0.06, which is obtained from maximizing revenue only, the profit is doubled.

In summary, the proposed Markov model is a fast and accurate method to estimate performance measures for parcel delivery services. It provides quick and suitable evaluations for each potential condition in the pricing problem. The golden section search algorithm properly analyzes the conditions and evaluations, and guides the search to the optimal solution. Based on the method proposed for the pricing problem, the parcel delivery services provider can easily decide the service price and improve profit.

5.2.5. Dynamic Pricing

The price mentioned in the previous discussion is fixed for all situations, regardless of the vehicle status and the number of customers waiting for the service. However, in practice, it is possible to quote different prices to customers in different situations. It will be interesting to find a pricing strategy that can quote different prices based on the vehicle status and the number of customers in the system in order to maximize overall profit.

It is assumed that the quoted price $x_{w,k}$ depends on the number of customers in the waiting list *w* and the number of delivery jobs left in the current vehicle trip *k*. According to the discrete choice model, the demand rate is as follows.

$$\lambda_{w,k} = \frac{\Lambda}{1+e^{U_0+\beta x_{w,k}}} \tag{5.7}$$

According to Equation (5.7), the demand rate varies in different situations. Providing the demand rate for every combination of *w* and *k*, the transportation costs and service levels can be obtained from the Markov model. The overall profit for parcel delivery services is calculated as Equation (5.8) based on a provided group of prices $x = \{x_{w,k}\}$.

$$f(x) = \sum_{w,k} x_{w,k}\lambda_{w,k}\sum_{I}\pi_{(w,k,I)} - \varphi_1 U - \varphi_2\sum_{w,k}(1-S_T)x_{w,k}\lambda_{w,k} \tag{5.8}$$

The golden section search is able to explore the solution space in only one dimension. However, the solution for $x_{w,k}$ has $w \times k$ dimensions. A conjugate gradient search is used to decide the suitable search direction for the golden section search. In each iteration, a direction r_n is calculated based on the direction r_{n-1} from the previous iteration. Initially, $r_0 = 0$.

$$r_n = d_n + \frac{\|d_n\|^2}{\|d_{n-1}\|^2} r_{n-1}$$

d_n is the gradient descent direction, obtained from the following equation.

$$d_n = \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

The best profit in direction r_n can be achieved at $x_n = x_{n-1} + \gamma_n r_n$, where γ_n is a probe length in one dimension obtained from the golden section search.

A set of experiments is carried out in order to find the dynamic pricing setting in a case with the parameters listed in Table 5.3.

Table 5.3.
Parameters of the dynamic pricing experiment

Size of region	Λ	T	φ_1	φ_2
100x100	1/40	300	0.3	1

"Λ" is the arrival rate of potential customers; "T" is the due time criterion; "φ_1" is the petrol cost per unit time when the vehicle travels on the road; "φ_2" is the penalty ratio compared to the price.

Table 5.4 shows the optimal results for each departure criterion. The optimal prices are not listed in the table, since they are a group of prices based on different situations. The prices in the case that the departure criterion is one are illustrated in Fig. 5.2. It shows that the prices are increasing with an increase of *w* or *k*. More customers tend to be attracted to make orders when fewer customers are in the system. On the contrary, customers tend to be pushed away when the vehicle is over-burdened. In the case of a 1/40 customer arrival rate, the system tends to stay in a state with small *w* and *k*, and has small probability in a state with *w* and *k* greater than five. Therefore, prices are more sensitive in the states with small *w* and *k*, and the surface of the graph is relatively flat in the state with large *w* and *k*.

Table 5.4.
Optimal solution for the dynamic pricing problem

Departure Criterion	Revenue	Utilization	Service Level	Profit	CPU Time
1	0.33082	0.667465	0.980963	0.129048	7843.54
2	0.33317	0.609092	0.944925	0.14578	7766.27
3	0.33466	0.586016	0.844370	0.14516	8662.77
4	0.32873	0.553190	0.664622	0.133984	14310.50
5	0.31464	0.501873	0.434248	0.117962	33921.19
6	0.29473	0.439273	0.211922	0.102888	31892.55

"Departure criterion" represents the minimum number of customers visited in a vehicle trip; "Revenue" represents the income of the service per unit time; "Utilization" represents the vehicle travelling cost; "Service Level" represents the percentage of delivery tasks completed within 300 units of time.

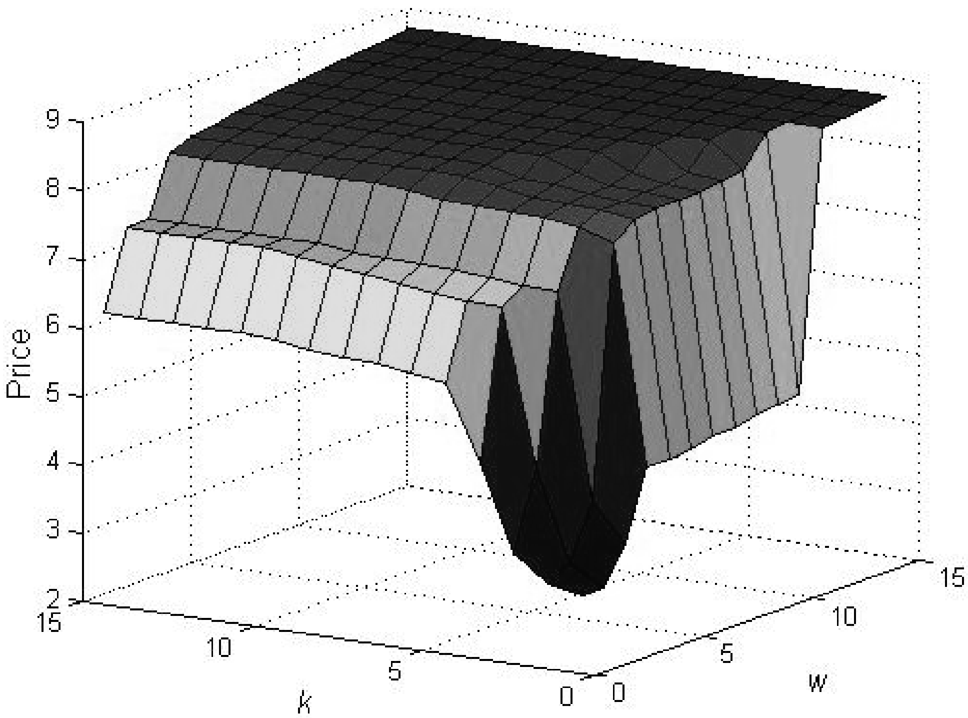


Fig. 5.2. Price setting based on *k* and *w*

The CPU time in Table 5.4 shows that the search algorithm for the dynamic pricing problem is not efficient. The solution from the combined algorithms of conjugate gradient search and golden section search may converge into a local optimum, and the final result may not be even close to the global optimal solution. Other optimization algorithms may efficiently generate better solutions for the pricing problem in parcel delivery services.

5.3. Network Design Problems for Parcel Delivery Services

5.3.1. Description of Network Design Problems

In the literature, there is a stream of research on the network design of transportation services. The network design problems are typical in the strategic planning of transportation systems, which is responsible for making long-term decisions on the physical structure concerning resources, locations and infrastructure (Wieberneit, 2008). The key issues in the network design of parcel transportation services include the location of warehouses, the assignment of service regions, and the scheduling and routing of vehicles.

Service providers are concerned with making strategic decisions of the following problems by a scientific investigation.

- If a completely new transportation system is designed, the number and locations of warehouses, and the service areas for the warehouses must be decided.
- It must be decided whether it is possible to close a warehouse in an existing network system to save operating costs and maintain the same service quality. In practice, this may occur if a parcel transportation company experiences a fall in demand in a particular region, or if advanced technology allows a neighboring warehouse to operate more economically in a larger area.

- It must be decided whether a new warehouse needs to be established in an existing parcel transportation system, and where the ideal location would be, if demand for the service increases dramatically or if the company wants to expand the business to a larger area.
- It must be decided whether cost savings can be achieved by changing warehouse locations or shifting service area boundaries of warehouses.

Researchers have proposed a number of models to address those concerns from parcel transportation service providers. Hwang (2002) divided the service region into sub-regions. A vehicle took care of the delivery to customers in each sub-region and the quality of service should be kept higher than a critical service level. The way to separate the service region and to decide the location of the distribution center was the main focus of Hwang's work. However, in this paper, the customers were fixed and known beforehand, and the final locations of warehouses were chosen from a discrete set of location sites. Bruns et al. (2000) designed the distribution system for Swiss Post. Swiss Post would like to choose the locations of delivery bases from a number of potential sites, from which parcels will be delivered. The area of Switzerland was divided into postal zones with each zone assigned to a delivery base. The delivery distance for a postal zone consisted of the stem distance and the running distance. The stem distance was proportional to the mean distance from the delivery bases to each customer in the postal zone. The running distance was calculated by the continuous approximation model (Langevin and Mbaraga, 1996). Wasner and Zapfel (2004) studied the Austrian parcel transportation networks. More than 2000 postal zones had to be assigned to depots and a hub. They considered adjusting service boundaries for existing depots, closing existing depots and setting up new depots. These papers discussed the warehouse locations and postal zone assignment in a vehicle routing circumstance. However, only transportation costs were taken as the optimization objectives in these papers, and the quality of service is ignored.

Pavone et al. (2011) suggested distributed algorithms for the partitioning problem based on the approximation of equitable power diagrams and equitable and median Voronoi diagrams. Their algorithms are applicable to resolve DVRP with multi vehicles based on the π-partitioning policy, in which the entire region is partitioned into m sub-regions and each vehicle is assigned to a sub-region executing the single-vehicle routing policy π to service demands that fall within the sub-region. They have claimed that their partitioning algorithms generate optimal solutions in heavy traffic situations and almost optimal solutions in light traffic situations. The key idea of their algorithms is to construct a suitable objective function, which reflects the weights of the generators in the Voronoi diagrams and the impact of the routing strategies. The setting of the objective function in general traffic situations is still worth for discussions. The Markov models proposed in this paper are effective methods to evaluate the impact of routing strategies on vehicle travel distances in a region and service levels. Based on the objective of reducing transportation cost and increasing service levels and profits, an optimization algorithm can be used to find a satisfactory design of the network in the service region.

5.3.2. Optimizing the Size of a Service Region

Before considering the complicated network design, deciding on a suitable size for a service region is discussed first. A business of parcel delivery services will be set up in a new city with one vehicle assigned to the service region. A customer is charged \$30.00 for a delivery. It is assumed that customers who require parcel delivery services at this price are uniformly distributed in the city with a density of $1/10^6$ per unit area. This means that customers appear with a Poisson distributed arrival rate $1/10^6$ within each unit of area in the region. The service provider promises that customer demands will be fulfilled within 300 units of time; otherwise, there will be 100% compensation for each overdue service. It is assumed that the vehicle travels at a constant speed and the petrol costs \$0.30 per unit time. The service provider has to decide the optimal size of each service region so that total profit is maximized.

In this case, the total demand rate is proportional to the area *A* of the service region, which is denoted as follows.

$$\lambda = A/10^6$$

The proposed Markov model is used to estimate the transportation cost and the overdue penalty for each value of *A*, based on the results from chapter 3. The profit can be calculated as follows.

$$f(A) = \frac{30A}{10^6} - 0.3U - 30\left(1 - S_T\right)\frac{A}{10^6} \tag{5.9}$$

The profit curve in terms of the size of the service region is illustrated in Fig. 5.3. The revenue, the travelling cost and the overdue penalty are monotonically increasing functions. Therefore, the profit function only has one peak above zero.

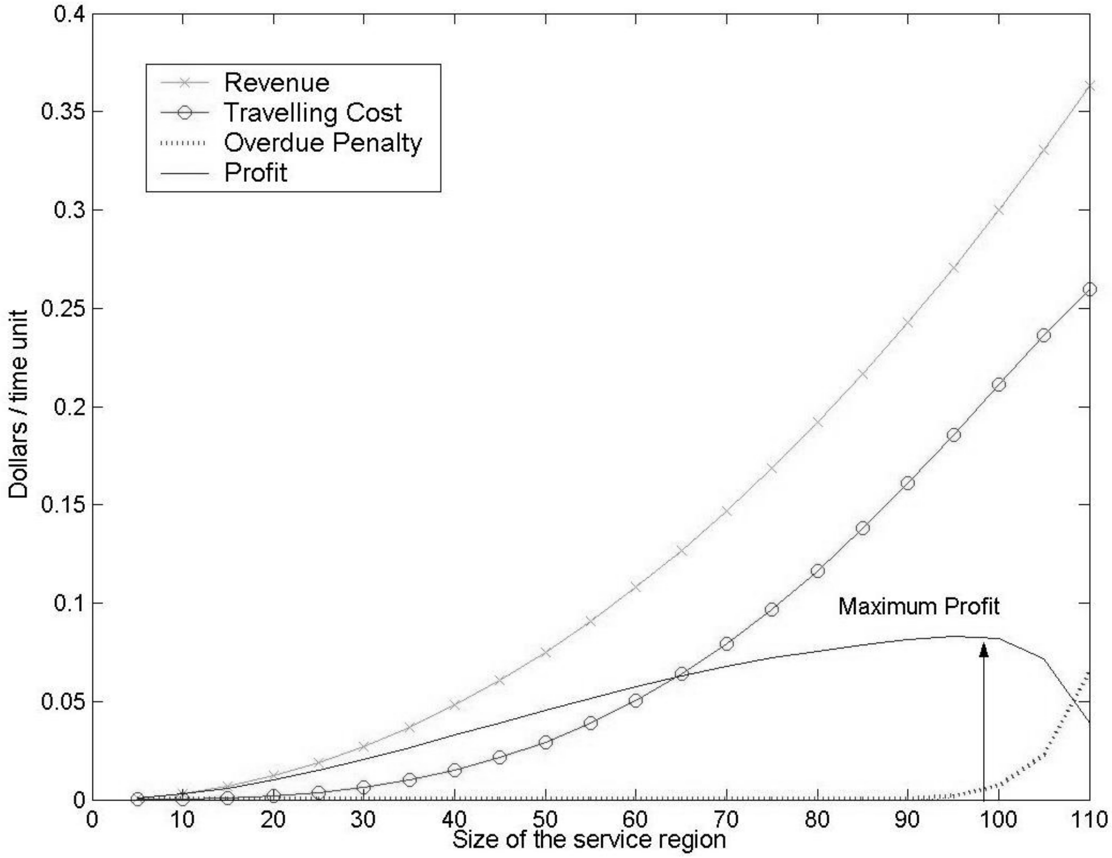


Fig. 5.3. Profit function in terms of the size of the service region

The golden section search mentioned in section 5.2.3 is applied to find the best region size. The decision variable must be replaced with the area of the service region. Furthermore, the search must be limited to $\lambda < 1/(0.6\sqrt{A})$, which means that $A < 14057$; otherwise, the demand will be far beyond the affordability of one vehicle. The golden section search is initiated with $A = 10^4$. The golden section search algorithm for the size optimization of a service region is as follows.

Step 1. Choose the initial area for the region. The decision variable is initiated with $A_1 = 10^4$. The one-step search length is *L*. Let $A_2 = A_1 + L$, $A_3 = A_1 + L/(1+\xi)$. Calculate the profit $f(A_1)$, $f(A_2)$ and $f(A_3)$ for the three cases.

Step 2. Decide the range of the search. If $f(A_1) > f(A_3) > f(A_2)$, let $L = \xi L$, $A_2 = A_3$, $A_3 = A_1$ and $A_1 = A_2 - L$. If $f(A_1) < f(A_3) < f(A_2)$, let $L = \xi L$, $A_1 = A_3$, $A_3 = A_2$ and $A_2 = A_1 + L$. Repeat step 2 until $f(A_3) \geq f(A_2), f(A_1)$.

Step 3. Search within the range $\Re = [A_1, A_2]$. Let $A_3 = A_1 + L/(1+\xi)$ and $A_4 = A_1 + \xi L/(1+\xi)$. If $f(A_3) > f(A_4)$, let $L = \xi L/(1+\xi)$. If $f(A_3) < f(A_4)$, let $L = \xi L/(1+\xi)$ and $A_1 = A_3$. Repeat step 3 until $L < \varepsilon$, where ε is a relatively small number.

Step 4. Finally, the optimal area is A_3 if $f(A_3)$ is greater, or A_4 if $f(A_4)$ is greater.

Based on the assumption that the service region is always a square in shape, the optimal solutions provided by the golden section search algorithm are shown in Table 5.5. It takes several minutes to achieve the optimality.

Table 5.5.
Optimal size of the service region

Departure Criterion	Area	Revenue	Utilization	Service Level	Profit	CPU Time
1	9277.38	0.278321	0.641373	0.990199	0.083181	240.46s
2	9331.10	0.279933	0.572584	0.943812	0.092429	296.40s
3	9141.87	0.274256	0.525600	0.828657	0.069584	373.98s
4	8193.63	0.245809	0.432591	0.644731	0.028703	445.78s

The symbols and abbreviations are the same as Table 5.1.

An optimization is carried out for each departure criterion. Thus, an optimal area for the service region is obtained for each one. For example, from the results in Table 5.5, the first row shows that it is better to operate parcel delivery services in a region with an area of 9277.38, when the vehicle always plans a trip with at least one customer. Actually, the service region is a 96.32x96.32 square. In this case, the maximum profit is \$0.083181 per unit time. In addition, it can be noted that the optimal area decreases as the departure criterion increases, when the departure criterion is greater than two. Due to the decreasing size of the service region, the demand within the entire service region decreases as well, which causes a reduction in revenue. The vehicle utilization also decreases, as the vehicle travels in a smaller area. Although demand decreases, the service level is still decreasing, since the vehicle spends more time waiting for customers at the warehouse due to an increase of the departure criterion. When the departure criterion is two, the profit is maximized in all cases. The profit starts decreasing when the departure criterion is greater than two, and it almost reach zero when the departure criterion is four. When the departure criterion exceeds four, the business starts losing money.

5.3.3. Region Partitioning

In this section, a similar optimizing problem with different decision variables is discussed. In this problem, a parcel delivery business is started in a new city. The size of the city is 10^4x10^4, which is too large to be served by one vehicle. The entire service region will be divided into a number of sub-regions. Each sub-region has a vehicle operating services and a warehouse located in the middle of the sub-region. Maintaining a warehouse will cost \$0.04 per unit time. The other conditions are the same as in section 5.3.2. A customer is charged \$30.00 for a delivery. The vehicle travelling cost is \$0.30 per unit time. The customer demand rate is $1/10^6$ per unit area. If the customer demand is not fulfilled within 300 units of time, 100% compensation will be provided. The objective is to find the optimal number of sub-regions that will maximize total profit.

To simplify the problem, it is assumed that each sub-region must be an identical square in shape. In this case, the demand rate within the sub-region is proportional to the area, which is denoted as $\lambda = A/N^2/10^6 = 100/N^2$, assuming that the entire region is divided into N^2 sub-regions.

Table 5.6.
Optimal number of sub-regions

N_D	N^2	Revenue	Utilization	Service Level	Profit / sub region	Total Profit	CPU Time
1	137x137	0.159838	0.298303	0.999963	0.070341	569.47406	126.41s
2	112x112	0.239158	0.456708	0.939842	0.087758	599.08278	164.56s
3	108x108	0.257202	0.478315	0.824152	0.068479	332.17565	211.92s
4	109x109	0.252504	0.450234	0.648088	0.028575	-135.7440	256.00s

"N_D" represents the minimum number of customers included in a vehicle trip. "N^2" represents the number of sub-regions. "Revenue" is the revenue gained in each sub-region. "Profit / sub region" represents the profit gained in each sub-region (excluding warehouse maintaining cost). "Total profit" is the overall profit gained from performing parcel delivery services in the city.

The optimal solutions are shown in Table 5.6 for different departure criterions (first column). The second column shows the optimal number of sub-regions. For example, with a departure criterion of one, the optimal number of sub-regions is 18769, which is denoted as 137x137 in the table. In each sub-region, the expected revenue is \$0.159838 per unit time. The utilization of each vehicle is estimated to be less than 30% and the service level is nearly 100%. With the optimal partitioning, the total profit is \$569.47406 per unit time for the entire service region.

The overall optimal profit, being \$599.08278 per unit time, is achieved with a departure criterion of two, when the entire service region is divided into 112x112 sub-regions. As the departure criterion increases, the optimal number of sub-regions decreases, which causes an increase of customer demand within each sub-region. Vehicle utilization is increasing, which indicates the vehicle is spending less time idle at the warehouse, even though the departure criterion is increasing, due to an increase of customer demands and an increase of the area covered by each vehicle. The service level sharply drops as expected, due to an increase of demand, service area and departure criterion. The profit dramatically decreases also, when the departure criterion is more than two, since the penalty for overdue services becomes the main factor in reducing profit. When the departure criterion is greater than four, the total profit is negative, since the profit earned in each sub-region is insufficient to pay for the fixed warehouse costs.

5.3.4. Network Design for Parcel Delivery Services

In this section, a transportation network is designed for a parcel delivery service. A parcel delivery business is started in a new city. The size of the city is 400x400, and the city is divided into 1600 square postal zones, each with an identical size of 10x10. Each postal zone has to be assigned to a nearby warehouse. Parcel delivery will be carried out by a vehicle belonging to the warehouse. The rental of each warehouse costs \$0.01 per unit time. The vehicle travelling cost is \$0.30 per unit time. The customer demands are uniformly distributed in the city with a density of $1/10^6$ per unit area, if a customer is charged \$30.00 for a delivery. If the customer demand is not fulfilled within 300 units of time, 100% compensation will be provided. A distribution center is located in the middle of the city. At the end of a service cycle consisting of 1000 units of time, a truck reaches a warehouse from the distribution center for replenishment. The travel time of the truck is proportional to the distance between the distribution center and the warehouse. The transportation costs of the truck are \$0.20 per unit time. It is assumed that no customer order will be rejected and products will never be out of stock due to the replenishment scheme. The objective is to find the suitable number of warehouses and the assignment of postal zones that will maximize total profit. To simplify the problem, it is assumed that the warehouse is always located at the geometric center of the sub-region.

All possible assignments of postal zones to warehouses comprise the solution space. A genetic algorithm (GA) is used to search the solution space to find satisfactory solutions for the network design problem. Postal zones are numbered from 1 to 1600 and warehouses are numbered from 1 to 100, as the company restricts the number of warehouses in the city to a maximum of 100.

The chromosome includes 1600 genes. The gene $g_j = i$ in the j^{th} position of the chromosome indicates that the postal zone *j* is assigned to the warehouse *i*. Some warehouses, not indicated in the chromosome, are not in charge of any postal zones and will be removed from the solution. Solutions will be evaluated by fitness values, which are the total profit f obtained within a service cycle (1000 units of time).

$$f = \sum_{i=1}^{N_v} f_i = \sum_{i=1}^{N_v} \left[\frac{30A_i}{10^3} S_{T,i} - 300U_i - 0.4D_i - 10 \right] \tag{5.10}$$

In Equation (5.10), N_v is the total number of warehouses operating delivery services. The total profit is obtained from the sum of the profit from every sub-region covered by a warehouse. f_i denotes as the profit from the sub-region of warehouse *i*.

$$f_i = \frac{30A_i}{10^3} S_{T,i} - 300U_i - 0.4D_i - 10 \tag{5.11}$$

The sub-region profit has four elements. The first element is the revenue generated from the successful fulfilled demand. A_i is the sum of postal zones areas covered by the i^{th} warehouse. Thus, the expected quantity of demand within a service cycle in the sub-region is $A_i/10^3$. $S_{T,i}$ is the service level representing the percentage of demands that are fulfilled within 300 units of time. The second element indicates the transportation costs of vehicle trips within the sub-region. U_i is the vehicle utilization representing the percentage of time the vehicle travels on the road. The third element is the transportation costs of the truck travelling between the warehouse and the distribution center. D_i is the travel time between the distribution center and the i^{th} warehouse. The rental cost of a warehouse within a service cycle is ten. U_i and $S_{T,i}$ are estimated by the proposed Markov model for every region. $\sum_{i=1}^{N_v} A_i = 400 \times 400$, since every postal zone has to be assigned to a warehouse.

The following steps show the scheme of the GA.

Step 1. INITALIZATION. Initialize a population *P* of size 20.

Step 2. SELECTION. Select a pair of individuals y, z from *P* as parents with regard to their fitness value.

Step 3. CROSSOVER. Generate a child z' applying the CROSSOVER operator to y and z.

Step 4. MUTATION1. Generate a modified child y' applying the MUTATION1 operator to y

Step 5. MUTATION2. Generate modified children y'' and z'' applying the MUTATION2 operator to y' and z', respectively, with a 50% chance.

Step 6. Insert y'' and z'' into *P* and in turn remove the two individuals with the worst fitness from *P*.

Step 7. Repeat Step 2 to 6 until the TERMINATION criterion is met.

Step 8. The satisfactory solution is the solution in the population *P* with the best fitness value.

In the INITIALIZATION step, the initial population of P is arbitrarily set. Postal zones belonging to the same warehouse must be clustered with each other by their locations in each individual of the initial population. For example, one of the individuals is composed of 25 warehouses equally scattered in the city (Fig. 5.4). Each warehouse is in charge of a sub-region with 8x8 postal zones. In Fig. 5.4, the number i on the sub-region indicates that the i^{th} warehouse is in charge of the postal zones within the sub-region.

1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

Fig. 5.4. An example of postal zone clustering

In the SELECTION step, y is randomly chosen from the population P. Each individual has a fitness value equal to the total profit from (5.10). The probability that one individual is chosen as y is proportional to its fitness value. However, z is chosen from the rest of the population with an equal chance.

The CROSSOVER operator is not easy to realize in this problem, since after CROSSOVER, it is usual to generate solutions with postal zones far away from each other that belong to the same sub-region. These assignments of postal zones are definitely unsatisfactory solutions and they cause difficulty in estimating the transportation costs and service level using the proposed Markov model. The GA selects five warehouses from individual z together with their postal zones, and replaces these warehouses and postal zones in individual y as new warehouses and sub-regions. Fig. 5.5 shows an example of the CROSSOVER. Postal zones belonging to the 6th warehouse in individual z is selected to replace the postal zones in y. In their child z', a new warehouse 26 is located where the 6th warehouse is located in z, and a new region is assigned to warehouse 26 with postal zones belonging to the 6th warehouse in z.

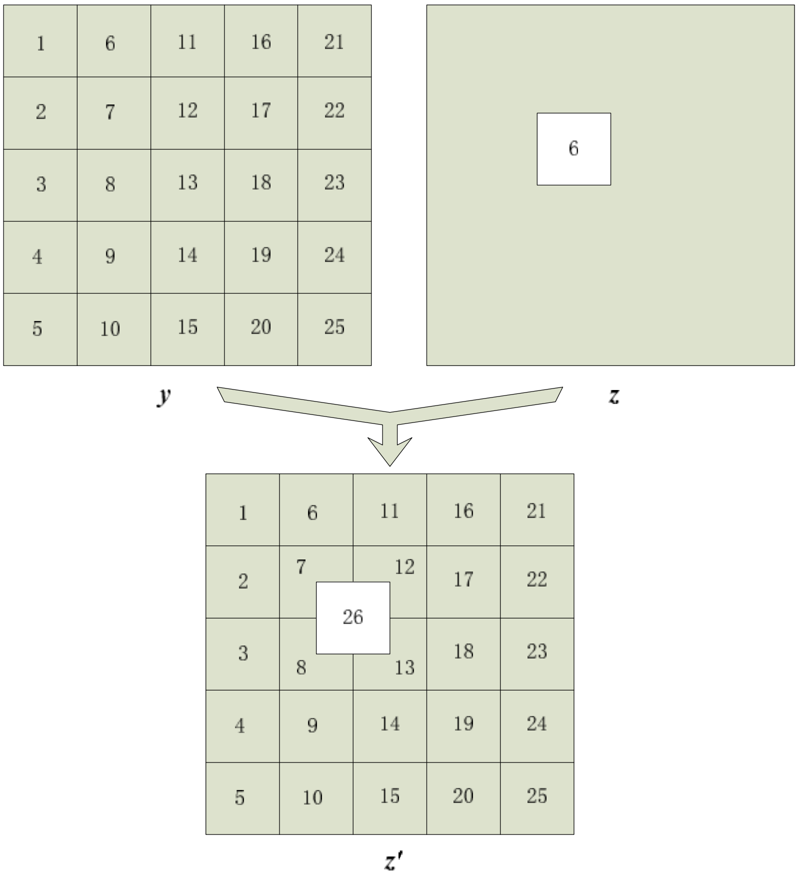


Fig. 5.5. An example of CROSSOVER

The MUTATION1 operator tries to extend the area of a sub-region covered by a selected warehouse. First, a warehouse is randomly selected. The warehouse with a lower sub-region profit f_i has a higher chance to be selected. The selected sub-region can be extended in a direction: (north, south, west, east, northwest, northeast, southwest or southeast), chosen at random with an equal chance. However, the region at the northwest corner of the city cannot be extended north, west, northwest, northeast or southwest. The postal zones adjacent to the selected sub-region in the selected direction are assigned to the selected warehouse. Fig. 5.6. shows an example of the MUTATION1 operator to warehouse 3. The region belonging to warehouse 3 is extended northeast.

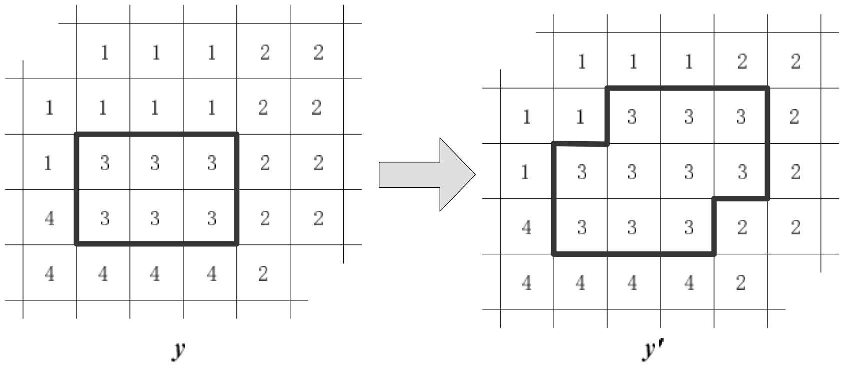


Fig. 5.6. An example of MUTATION1

The MUTATION2 operator tries to reassign a postal zone to another warehouse nearby. The selected postal zone must be located at the edge of a sub-region, which means that at least one of the postal zones nearby belongs to another warehouse. Fig. 5.7 is an example of MUTATION2. The number located at the lower right corner of a postal zone indicates the number of the postal zone. Postal zone 93 belonging to warehouse 3 in individual y' is chosen for MUTATION2. There are two postal zones nearby belonging to warehouse 1 and 2. Postal zone 93 has an equal chance to be assigned to warehouse 1 or 2 after MUTATION2.

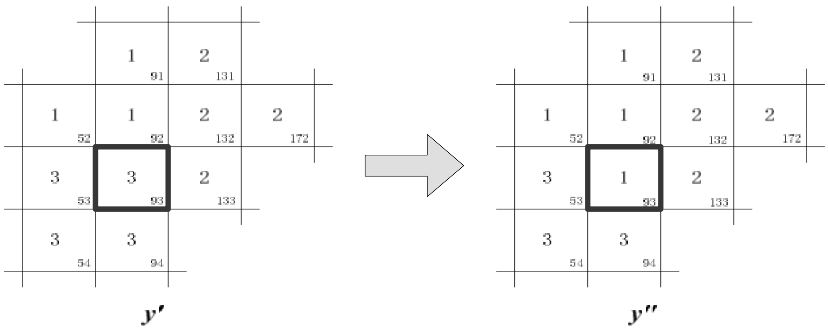


Fig. 5.7. An example of MUTATION2

Finally, the TERMINATION criterion is defined. The search terminates after 50 iterations without an improvement of the population *P*. The fitness values of children y'' and z'' are estimated. If all of them are worse than the value of the third from the bottom in *P* for 50 times in a row, the GA is terminated.

Initially, the total profit for the postal zone assignment in Fig.5.4 is \$188.33. Actually, the initial solution in Fig. 5.4 is a good solution, close to the best solution that is eventually found. After optimization using the GA, the city is still divided into 25 regions. The total profit for the best found solution is \$233.38. The region's profit ranges from \$-14.46 to \$48.59. The regions with the least profit are located at the four corners of the city, while the regions with the most profit are located at the centre of the city. In summary, the GA significantly improves the transportation network of delivery services, since the proposed Markov model provides a fast and accurate estimation of transportation costs and service levels.

The solution obtained using GA can be further improved by applying better MUTATION and CROSSOVER operators. Furthermore, the location of the warehouses can be optimized further. They do not necessarily have to be located at the geometric centres of the sub-region. There may be more suitable locations between the geometric centres of the sub-region and the distribution centre.

5.4. Order Acceptance Problem

5.4.1. Description of the Order Acceptance Problem

Order acceptance is a tactical managerial activity that deals with accepting and rejecting customer orders. It may also deal with related decisions, such as due time quotations. From the perspective of a professional practice, a firm may choose to reject potential orders for market focus, competitive advantage or capacity limitation reasons. When an overabundance of orders is encountered by manufacturing or service companies, such that demand exceeds capacity, some difficult choices are brought upon the companies. There is an important trade-off between the profit associated with an order and the cost of capacity that may be diverted from others. In addition, late deliveries may result in penalties of long-term loss of goodwill and market share. In a competitive market, the importance of on-time delivery may make it cost- and profit-effective to reject some orders.

In the literature, an order in the manufacturing or service industry is usually modelled together with revenue and the time that workers need to spend on it. A decision needs to be made on which orders to accept in order to earn the most profit, while the total time spent on the accepted orders are within a limit. In most cases, the revenue and time associated with an order are independent of other orders. However, for transportation services, the cost and time spent on each order tightly depend on other orders. This makes the problem even more difficult to analyze. In the dynamic vehicle routing problem with time windows, a time window associated with each customer specifies when the transportation service has to be completed by. The routing strategies try to add new demands into the existing schedule (Bent and Van Hentenryck 2004; Coslovich et al. 2006). If the service for a new demand exceeds the time window, or adding the new demand into the existing schedule will cause other demands to violate their time windows, the new demand will be rejected. This is the usual application of the order acceptance in the DVRP. All papers in this research area focus on the routing strategies instead of order acceptance itself.

In this section, the order acceptance problem in parcel delivery services will be analyzed in another way. Sometimes, logistics companies or supermarkets guarantee delivery within a certain number of days or hours. For example, DHL charges different delivery fees for same day, next day or a week later deliveries. Carrefour offers same day delivery between 4:00pm and 7:00pm if orders are made before 12:00pm, and same day delivery between 7:00pm and 9:00pm if orders are made between 12:00pm and 4:00pm. Pizza Hut can estimate the delivery time according to its workload. Customers usually have expectations on the waiting time. If the estimated delivery time is much longer than their expectations, they will refuse to make orders. Therefore, deciding on a guaranteed delivery time is a challenging task for service providers. A large estimated delivery time may cause the company to lose customers, while a short delivery time may increase the burden on the delivery service and harm to the service level. By determining a suitable delivery time, service providers can focus on the group of customers who may be satisfied with the service and are willing to pay for it.

The problem discussed in this section is about a new parcel delivery business in a city. The service provider plans to assign a vehicle to deliver products within a 100x100 square region and is going to charge \$30.00 for each delivery. The provider also announces that the service can be completed within an estimated delivery time T_D. The vehicle travelling cost is \$0.30 per unit time. The customer arrival rate to check for the information on the delivery service is Λ, and the acceptable waiting time for each customer is assumed to follow a uniform distribution $U(100, 600)$ with a minimum value of 100 units of time and a maximum value of 600 units of time. Customers with an acceptable waiting time longer than T_D are willing to make a delivery order. If the order is not fulfilled within the estimated delivery time, \$30.00 in compensation will be provided. The service provider cannot reject any order made by a customer. The only thing they can do is to adjust the estimated delivery time, and let customers make their choices. Therefore, the objective is to find the estimated delivery time T_D that will maximize total profit.

Since the acceptable waiting time follows a uniform distribution $U(100,600)$, the percentage of customers who will accept the estimated delivery time is $(600-T_D)/500$, and the actual demand rate for the delivery service is as follows.

$$\lambda = \frac{(600 - T_D)\Lambda}{500}$$

The proposed Markov model in chapter 3 is used to estimate the transportation costs and the service level for each value of T_D. The profit can be calculated as follows.

$$f(T_D) = \frac{30(600 - T_D)\Lambda}{500} - 0.3U - 30\left(1 - S_T\right)\frac{(600 - T_D)\Lambda}{500} \tag{5.12}$$

In the case of $\Lambda = 1/40$, the profit curve in terms of the estimated delivery time is illustrated in Fig. 5.8. The revenue, travelling cost and the overdue penalty functions are monotonically decreasing. The profit function has only one peak above zero, when T_D is near 355 units of time. The golden section search mentioned in section 5.2.3 can be applied to find the best-estimated delivery time. The decision variable must be replaced with the estimated delivery time T_D. Furthermore, the search is limited to $T_D \geq 200$; otherwise, the demand will exceed what one vehicle can handle. The golden section search is initiated with $T_D = 300$.

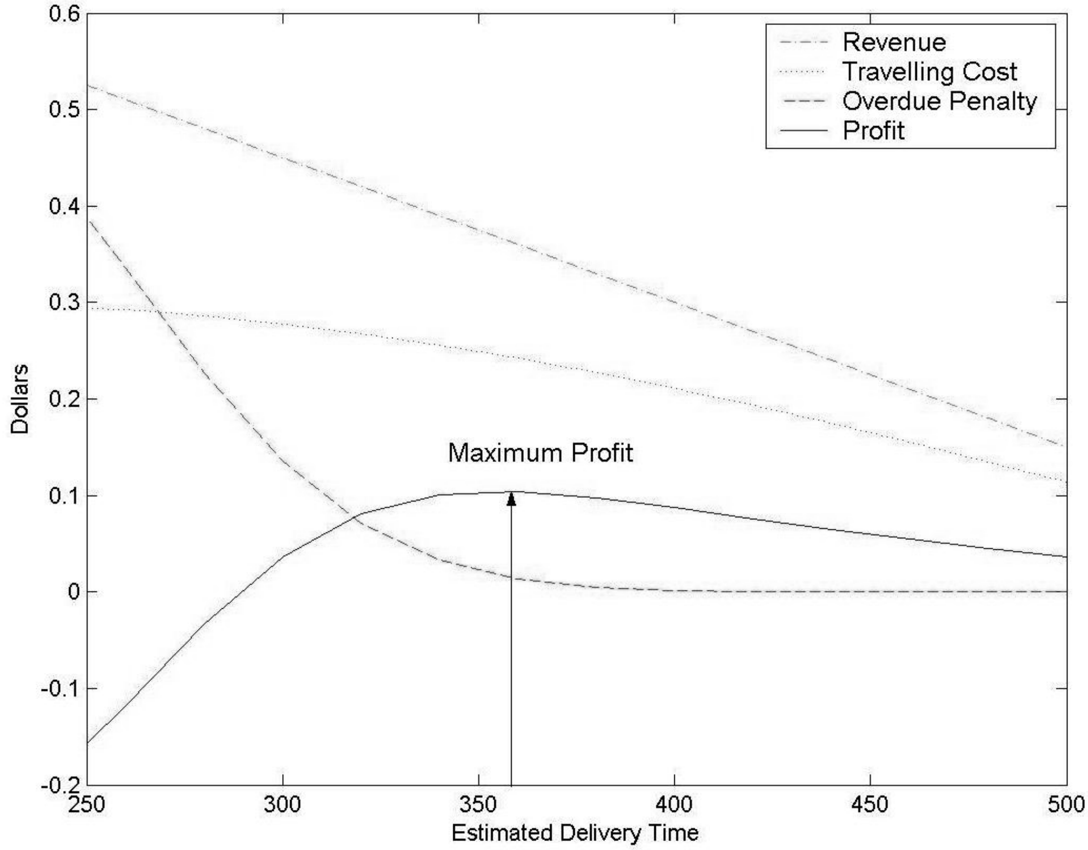


Fig. 5.8. Profit function in terms of the estimated delivery time ($\Lambda = 1/40$)

5.4.2. Optimization Results and Discussions

In the first set of optimization experiments, the golden section search is applied to find the best solution for problems with different customer arrival rates $\Lambda = \{1/20, 1/40, 1/60, 1/80\}$. The results are shown in Table 5.7.

The first column in Table 5.7 indicates the arrival rates of potential customers looking for delivery services. The second column lists the optimal solutions for the estimated delivery time, which can be provided as a quotation to potential customers. For example, the optimal estimated delivery time for the case of a 1/20 customer arrival rate is 462.023 units of time. After reviewing a quotation of the estimated delivery time, a number of potential customers decide to make orders. Fulfilling these orders provides revenue of \$0.41393 per unit time. By setting the estimated delivery time at 462.023, the service provider can focus on a group of potential customers and achieve a service level above 95% with vehicle utilization above 88.5%. Finally, the service provider can earn a maximum profit of $0.127914 per unit time.

Table 5.7.
Optimal solution of estimated delivery time ($N_D = 1$)

Λ	T_D	Revenue	Utilization	Service Level	Profit	CPU Time
1/20	462.023	0.413930	0.885053	0.950348	0.127914	454.78s
1/40	354.587	0.368119	0.820454	0.952110	0.104416	288.22s
1/60	284.112	0.315888	0.732784	0.958885	0.083105	178.95s
1/80	242.562	0.268079	0.641131	0.966505	0.066784	125.70s

" Λ " represents the customer arrival rate to check for information on the delivery service.
" T_D " represents the estimated delivery time provided by the service provider.

As the customer arrival rate decreases, less potential customers require delivery services. Thus, the estimated delivery time is set to a lower value to attract more customers. The revenue is still decreasing, the vehicle is idle longer at the warehouse, the service level is improving, and the final profit is decreasing. It can be noted that the vehicle is capable of managing all delivery requests from potential customers in the case of 1/60 and 1/80 customer arrival rates. However, the optimal solution shows that some customers with high expectations still need to be removed from the scope of the services. The customers with high expectations could be served by special express delivery services with additional charges.

In the second set of optimization experiments, the best solutions are found for problems with various vehicle departure criterions. The potential customer arrival rate is set at 1/40 in the experiments. The vehicle can start a new trip with at least one to six pending customer demands. The results are shown in Table 5.8.

Table 5.8.
Optimal solution of estimated delivery time ($\Lambda = 1/40$)

N_D	T_D	Revenue	Utilization	Service Level	Profit	CPU Time
1	354.587	0.368119	0.820454	0.952110	0.104416	288.22s
2	362.742	0.355887	0.736018	0.947658	0.116511	305.36s
3	371.278	0.343084	0.680561	0.909460	0.107900	337.66s
4	383.180	0.325230	0.628573	0.825428	0.079925	389.69s
5	400.479	0.299282	0.568210	0.695868	0.037829	429.09s
6	427.944	0.258084	0.483638	0.539205	-0.005908	474.16s

" N_D " represents the minimum number of customers included in a vehicle trip. " T_D " represents the estimated delivery time provided by the service provider.

Different departure criterion is listed in the first column in Table 5.8. The second column shows the optimal solutions for the estimated delivery time. For example, the optimal estimated delivery time for the case with a departure criterion of three is 371.278 units of time. With this quotation for the delivery time, the corresponding expected revenue is \$0.343084 per unit time, and the final profit is \$0.1079 per unit time with vehicle utilization over 68% and a service level over 90%.

The overall optimal profit, which is \$0.116511 per unit time, is achieved in the case with a departure criterion of two, when the estimated delivery time is set to 362.742 units of time. As the departure criterion increases, the optimal estimated delivery time increases, which removes more customers from the scope of the services. Thus, the revenue decreases further. The vehicle utilization also decreases, since there are fewer orders for the delivery service and the vehicle trip becomes less efficient with fewer customers. The service level sharply decreases, as expected, since the vehicle spends more time idle at the warehouse waiting for enough customers to start a delivery trip. The profit increases at first, as there is efficiency in the vehicle trip is achieved. It decreases when the departure criterion is more than two, because the penalty for overdue services becomes the main factor in reducing profit. When the departure criterion is greater than six, the total profit becomes negative, since it is difficult to keep a high service level and keep customer satisfied in this situation.

In summary, the proposed Markov model provides quick and accurate evaluations for each condition proposed by the order acceptance problem. The golden section search algorithm properly analyzes the conditions and evaluations, and provides the optimal solution. Based on the method proposed for the order acceptance problem, the provider of parcel delivery services can easily determine the estimated delivery time and improve profit.

6. Conclusions and Future Research

6.1. Conclusions

In this thesis, parcel transportation services were investigated. In such services, small packages are transported from or to customers. The service providers are required to coordinate the transportation effectively and efficiently in order to meet the transportation needs of customers Especially in the e-commerce business, quality and reliability of parcel transportation services is of utmost importance as the product has to be delivered on time and in perfect shape at the customer location. A reliable parcel transportation service is, in fact, the backbone of an Ecommerce business. However, due to the complicity of the routes and schedules of vehicles in parcel transportation services, it is difficult to evaluate the cost, the service level and other performance measurements.

In the literature, a few routing strategies have been published to find out better schedules for specific parcel delivery cases. However, seldom researchers devoted to the evaluation of the performance of the service. An efficient and systematical method to estimate various performance measures, such as transportation cost and service levels, for the service is still lacking.

This thesis provides a systematical method to evaluate the parcel delivery service. The stochastic model derived is an efficient tool to estimate transportation cost and service levels for the service in various situations. Firstly, although this model cannot provide a detail schedule of vehicle traveling around customers for a specific case, it can evaluate the impact to cost and service levels when different routing strategies are applied. Secondly, this model can be used to evaluate whether a logistics system is optimized. Potential improvements for the logistics system could be found based on the evaluation.

Corresponding management decision about improvements can then be made. Thirdly, the model may help to answer the following questions which may be interesting to logistics managers. Although literature few other groups have tried to answer these questions in different scenarios, seldom of them is solving the problems under the e-business and dynamic parcel delivery circumstance, which involves dynamic demand and vehicle routing.

1. What kind of promise about the delivery time can be made to customers based on a demand distribution, which may not decrease the intension of buying the service and make a profit?
2. What kind of pricing rule can be used to maximize the profit?
3. How to choose a suitable location for depots and how to assign customers to depots to minimize delivery cost and maximize service level?
4. Selection of vehicle types, the number of vehicles rent for the service, evaluation of the performance of the third party carrier in the last mile delivery, and differentiation of demands for prime services.

The detail contribution in each chapter is elaborated as follows. In chapter 3, a novel Markov model was proposed to estimate transportation costs and service levels. Based on the estimation of means and variances of vehicle travel time proposed by Christofides and Eilon (1969) and Larson and Odoni (1981), a hypo-exponentially distributed vehicle travel time was applied to the Markov model. The number of customers in the system and the status of the vehicle were defined as the system state. The steady state process of the Markov model could be applied to analyze vehicle utilization, which could then be used to calculate transportation costs. A customer waiting process of the Markov model tracking the entire process of a specific customer in the system was used to measure the customer waiting time and to calculate the service levels for parcel transportation services. Experiments have demonstrated that the proposed Markov model was able to provide accurate estimations of various performance measures for parcel transportation services with less than 10% errors compared to results from discrete event simulation, and it is 100 times faster. Especially in the cases that vehicle utilization is between 50% and 90%, the errors are only about 5%.

The derived model is flexible and is capable to solve variety of extension of the parcel transportation service. Four additional constraints and issues that could be encountered in parcel transportation services were discussed in chapter 4.

- The capacity constraint of the vehicle was analyzed. A package ordered by a customer was assumed to only occupy one unit of capacity in the vehicle. Thus, the capacity constraint was converted to the limit of customers visited by the vehicle on a single trip. A modified version of the Markov model was used to incorporate the characteristics of this specific constraint. The modified Markov model was able to provide reliable estimations of the performance measures for parcel delivery services with capacity constraints and its outputs were consistent with the experimental results with differences less than 5%. Furthermore, the versatility of the proposed approach had been demonstrated in the selection of vehicles with various capacities and petrol consumptions.
- A multiple vehicle issue was also analyzed. To simplify the problem, the vehicles were assumed to be identical. The number of customers in the system and the status of one of the vehicles were defined as the system state. It was further

assumed that the status of the observed vehicle was independent of the status of the other vehicles. The modified Markov model in this section is capable to estimate performance measures with acceptable accuracy. The errors of average customer waiting time are less than 20%. These errors may mainly come from the assumptions which ignore the dependency between states of vehicles. Furthermore, the model could be applied in the decision-making of the number of vehicles operating in a service region.

- Dynamic delivery services and dynamic pickup services were compared against each other. In dynamic delivery services, the vehicle trip starts off from a warehouse, travels to a number of customers and ends up at some warehouse. On the other hand, the warehouse is ignored in dynamic pickup services, and the vehicle is always travelling from one customer to another. This is the main difference between the two kinds of transportation services. Different Markov models were developed for these two kinds of services, respectively. Experiments have demonstrated that the Markov model for dynamic pickup services has less than 10% errors compared to the discrete event simulations, which is less accurate than the model for dynamic delivery services.
- The routing strategies in parcel transportation services were also discussed. Exact algorithms and heuristics tend to generate more efficient vehicle routing schedules. Two routing algorithms were investigated: a Branch-and-Bound algorithm and a best-insertion algorithm. A hyper-hypo-exponential distribution was used to approximate the vehicle travel time on the route generated by the two algorithms. A modified Markov model was applied to estimate the performance

measures. Experiment show that the hyper-hypo-exponential distribution actually does not well approximate the distribution of vehicle travel time. However, the results generated by the Markov model do not significantly rely on the accuracy of the estimation of the vehicle travel time. The final results of transportation cost and customer waiting time have less than 10% errors compared to that from discrete event simulations, which demonstrated that the modified Markov model was able to generate fast and accurate estimations for parcel transportation services with various routing strategies. Furthermore, the modified model derived in this section can be used to evaluate whether a routing strategy is better than the other, and whether a strategy is better in reducing cost or improving service levels.

In the real-life application, managers are able to evaluate the performance of their business efficiently with the proposed Markov model, and find out potential improvements to increase the profit based on the estimation of performance measures. Three management decisions which can be made with the assistance of the proposed Markov model were discussed in chapter 5.

- A pricing problem. The relationship between the service price and customer demands was assumed to satisfy the discrete choice model. If a certain price was provided, the proposed Markov model was able to provide estimations of transportation costs, service levels and profit. A golden section search algorithm and a conjugate gradient search were applied to find the best price in order to maximize total profit. Experiments have demonstrated that the proposed model for this pricing problem has less than 5% errors compared to discrete event simulations, and it is 20 times faster. Based on the proposed model, it is able to find optimal solutions in minutes for the fixed pricing problem, and to find satisfactory solutions in hours for the dynamic pricing problem. The best price decided by the proposed model may double the profits compared to that from an algorithm in order to optimize sales. A group of optimized results showed that the Markov model was able to assist service providers in deciding the best price for parcel delivery services and in improving their business.

- The network design problem. The objective was to design a distribution network within a city. The service provider had to decide the number of warehouses and the warehouse locations. In addition, every postal zone in the city had to be assigned to a warehouse. A golden section search was applied to determine the best size for a service region and the best way to partition the entire city in minutes. A genetic algorithm was used to find a satisfactory assignment of postal zones to warehouses. The proposed method successfully generated good solutions that significantly improved profit.

An order acceptance problem. In this scenario, an estimated delivery time is provided to potential customers before they book parcel transportation services. If the estimated delivery time is beyond the acceptable customer waiting time, the service provider will lose the potential order. A golden section search based on the performance measures estimated by the Markov model takes several minutes to provide optimal solutions of the estimated delivery time in order to maximize profit.

In summary, the proposed Markov model was able to analyze parcel transportation services in a systematic manner and was able to generate fast and accurate estimations of performance measures. It is a flexible model and can be modified to adapt to new constraints and circumstances. Furthermore, it is a particularly useful tool that can assist service providers in making decisions on various real-life situations.

6.2. Future Research Perspectives

In this thesis, the Markov model was applied to address dynamic parcel transportation services to potentially improve the efficiency in several decision-making issues. The proposed Markov models provided the mathematical tools to estimate performance measures and resolved realistic problems. In section 6.2.1, further improvement on the accuracy of the Markov models is suggested using several potential methodologies. In addition, other research problems relevant to parcel transportation services remain unexplored. From sections 6.2.2 to 6.2.4, several promising problems are highlighted for further research.

6.2.1. Further Improvement of the Markov Models

- When the demand rate is relatively large in heavy traffic conditions, the accuracy of the estimation of vehicle utilization and customer waiting time by the proposed Markov model is slightly lower. This is because the number of customers queued for services tends to be larger in heavy traffic conditions, and there is a higher chance the default queue length in the Markov model could be exceeded. Extending the queue length will definitely improve the results. However, it will significantly expand the state space and lead to a heavy computational burden. Other methods or models need to be considered to reduce the state space and increase the efficiency of the solution. It might be worth trying the continuous state Markov process and the diffusion approximation model, which were applied in the analysis of the G/G/m queuing system (Newell, 1973; Kleinrock, 1976;

Halachmi and Franta, 1978; Sunaga et al. 1978; and Kimura, 1983).

- In section 5.4, an order acceptance problem was analyzed for parcel delivery services. An estimation of delivery time is provided to potential customers. They evaluate whether the estimated delivery time is within their expectation and decide whether to make orders. This estimated delivery time in section 5.4 was applied to all situations, regardless of the vehicle status and the number of customers in the system. However, it would be wise to provide a different estimated delivery time based on the status of the vehicle and the queue in order to maximize overall profit. For example, when there are fewer customers in the system, the estimated delivery time could be shortened to attract more customers.

It is assumed that the estimated delivery time $T_{D,w,k}$ depends on the number of customers in the waiting list w and the number of delivery tasks left in the current vehicle trip k. If it is assumed that the acceptable customer waiting time follows a uniform distribution $U(100,600)$, the demand rate is as follows.

$$\lambda_{w,k} = \frac{(600 - T_{D,w,k})\Lambda}{500} \tag{6.1}$$

According to Equation (6.1), the demand rate varies in different situations. Providing the demand rate for every situation, the transportation cost, service level and total profit can be calculated by the Markov model. An optimization algorithm can be applied to search for the optimal solution in the order acceptance problem.

6.2.2. Parcel Delivery Services with Finite Products Stored in the Warehouse

In this thesis, the products stored in the warehouse were assumed to be infinite. However, this assumption may not always be satisfied in practice. Since supply delay and shortage of products may significantly increase the delivery time and decrease the service level, the proposed Markov model in this thesis will need to be modified.

If it can be assumed that products become out of stock with equal probability at any time, the model can be simply modified by reducing the transition rate out of the state in which the vehicle is idle at the warehouse. However, the inventory in the warehouse is usually monitored using the (s,Q) policy in practice. Under this policy, if the inventory is equal to or less than a replenishment point s, the supplier will subsequently deliver a quantity of Q products to the warehouse. The integration of the product supply to the warehouse and delivery to customers can be modeled as a queuing network. The decomposition method proposed by Chen (2010) incorporated with the proposed Markov model in this thesis may provide estimations of performance measures. Based on the performance measures, optimization algorithms may be used to determine an appropriate replenishment point and quantity.

6.2.3. Dynamic Traffic Conditions

In this thesis, dynamic customer demands were the main consideration in the study of parcel transportation services, and the analysis of dynamic traffic conditions was lacking. However, travel time in urban areas fluctuates due to a variety of factors, such as accidents, traffic conditions and weather conditions (Haghani and Jung, 2005). Ignoring these travel time variations can result in inaccurate estimations of transportation costs, customer waiting times and service levels. Due to dynamic traffic conditions, the time that the vehicle spends on the road becomes more unpredictable. Furthermore, the vehicle may occasionally break down, and the rest of the transportation tasks have to be postponed until the vehicle is fixed. In these circumstances, it is more difficult to estimate performance measures for parcel transportation services.

In this thesis, the travel distances between customers and the warehouse have been modelled as a random variable, due to uncertainty regarding customers' locations. In dynamic traffic conditions, the velocity of the vehicle can be modelled as another random variable. The vehicle travel time is calculated as the ratio between the travel distance and the vehicle speed. The mean and variance of the vehicle travel time can be calculated accordingly based on the distributions of the travel distance and the vehicle speed. Based on the means and variances of travel time, the proposed Markov model can be used to estimate performance measures in dynamic traffic conditions.

When the occasional event of a vehicle breaking down is considered, it has to be assumed that the vehicle breaks down according to a Poisson process and the repair time follows an exponential distribution. By adding a Poisson process for the vehicle breaking down and an exponential distribution for the repair time, the situation of vehicle breaking down can be monitored. The system state needs one more variable to indicate whether the vehicle is "working" or "broken". The transition rate between a "working" state and a "broken" state is the vehicle breakdown rate and repair rate. Further modified Markov models can be built to estimate performance measures in this case.

6.2.4. Dynamic Dial-A-Ride Systems

In dial-a-ride transportation systems, goods must be transported from an origin to a destination. In this problem, the vehicle has to visit two locations for a request. Taxi and home moving services are two typical examples of dial-a-ride transportation systems. A number of North American and European countries also provide public dial-a-ride transportation services for elderly and handicapped persons. Customers normally call in to book for services, and vehicle scheduling and routing must be generated to fulfil demands. As shown in Fig. 6.1, after receiving an assigned task, the vehicle follows the routing instruction to pick up the required goods from the origin and bring them to the destination. After that, the vehicle will relocate its position, preparing for the next request if it does not have any further assignments.

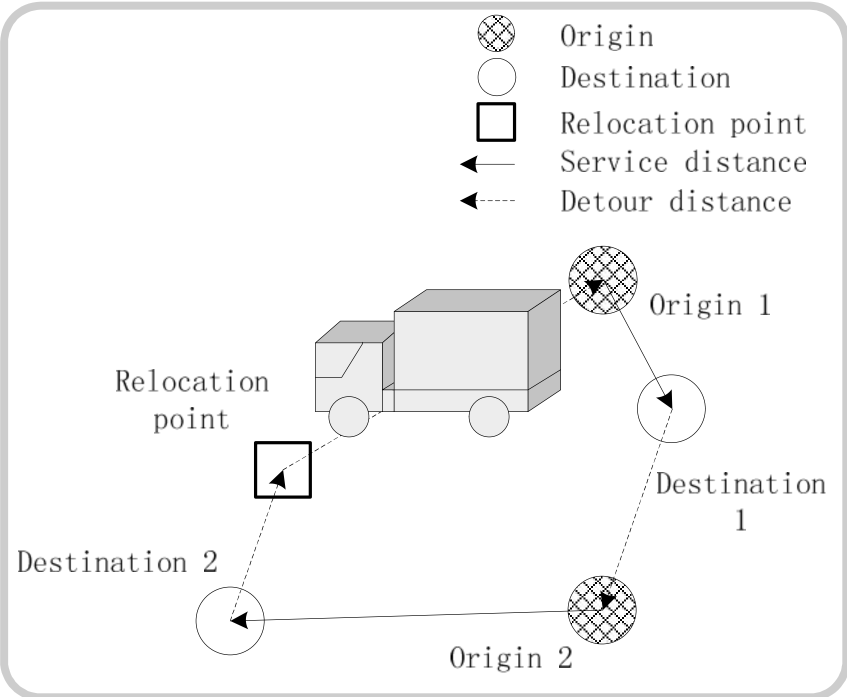


Fig. 6.1. A vehicle performing dial-a-ride transportation services

A number of vehicle routing and demand assignment strategies have been proposed in the literature to minimize transportation costs (Attanasio et al. 2004) and travel distance when vehicles are empty (Teodorovic and Radivojevic, 2000). However, no mathematical model has been proposed to estimate various performance measures for dial-a-ride services.

The proposed Markov model in this thesis can be modified to evaluate dial-a-ride transportation systems. One more variable can be used to indicate whether the vehicle is heading to an origin or a destination. In order to fulfil each dial-a-ride service request, the vehicle has to travel through two paths. One path is towards the origin followed by the other path towards the destination. Performance measures can be estimated using the similar Markov process mentioned in Chapter 3. Based on the estimation of various performance measures, the model can be used to analyze the issue of vehicle relocation after fulfilling a customer's demand and to assist service providers in making some management decisions, such as the service price and the number of vehicles in a service region.

Bibliography

- Abad, P.L. and C.K. Jaggi. A Joint Approach for Setting Unit Price and The Length of The Credit Period For a Seller When End Demand is Price Sensitive, Internal Journal of Production Economics, 83, pp.115-122. 2003.
- Ahmmed, A., M.A.A. Rana, A.A.M. Mahmudul Haque and M. Al Mamun. A Multiple Ant Colony System for Dynamic Vehicle Routing Problem with Time Window. ICCIT, vol. 2, In 2008 Third International Conference on Convergence and Hybrid Information Technology, November 2008, Busan, South Korea, pp.182-187.
- Amberg, A., W. Domschke and S. Voβ. Multiple center capacitated arc routing problems: A tabu search algorithm using capacitated trees, European Journal of Operational Research, 124, pp.360-376. 2000.
- Angelelli, E., N. Bianchessi, R. Mansini, M.G. Speranza. Short Term Strategies for a Dynamic Multi-Period Routing Problem, Transportation Research Part C, 17, pp.106-119. 2009.
- Asdemir, K., V.S. Jacob and R. Krishnan. Dynamic Pricing of Multiple Home Delivery Options, European Journal of Operational research, 196, pp.246-257. 2009.
- Attanasio, A. J-F. Cordeau, G. Ghiani and G. Laporte. Parallel tabu search heuristics for the dynamic multi-vehicle dial-a-ride problem, Parallel Computing, v30, n3, pp.377-387. 2004.

- Ausiello, G., E. Feuerstein, S. Leonardi, L. Stougie and M. Talamo. Serving requests with on-line routing. In Proc. 4th SWAT' 94, Lecture Notes in Computer Science, Vol. 824, Springer, Berlin, pp. 37–48, 1994.
- Ausiello, G., E. Feuerstein, S. Leonardi, L. Stougie and M. Talamo. Competitive algorithms for the on-line traveling salesman. In Algorithms and Data Structures. 4th International Workshop, WADS '95, Kingston, Canada, August 1995, pp.206-217.
- Ausiello, G., E. Feuerstein, S. Leonardi, L. Stougie and M. Talamo. Algorithms for the on-line traveling salesman, Algorithmica, 29, pp. 560–581. 2001.
- Barbucha, D. and P. Jedrzejowicz. Multi-agent Platform for Solving the Dynamic Vehicle Routing Problem. In Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems, Beijing, China, October 12-15 2008, pp.517-522.
- Bent, R.W. and P. Van Hentenryck. Scenario-Based Planning for Partially Dynamic Vehicle Routing with Stochastic Customers, Operations Research, v52, n6, pp.977-987. 2004.
- Beardwood, J., J. H. Halton and J. M. Hammersley. The Shortest Path Through Many Points, Proceedings of Cambridge Philosophical Society, v55, i4, pp.299-328. 1959.
- Bertsimas, D.J. and G. Ryzin. A Stochastic and Dynamic Vehicle Routing Problem in the Euclidean Plane, Operations Research, Vol. 39, No. 4., pp. 601-615. 1991.

- Bertsimas, D.J. and G.J. van Ryzin. Stochastic and dynamic vehicle routing in the Euclidean plane with multiple capacitated vehicles, Operations Research, Vol 41, pp. 60-76. 1993a.
- Bertsimas, D.J. and G.J. van Ryzin. Stochastic and dynamic vehicle routing with general interarrival and service time distributions, Advances in Applied Probability, Vol. 25, pp. 947-978. 1993b.
- Bianchessi, N. and G. Righini. Heuristic algorithms for the vehicle routing problem with simultaneous pick-up and delivery, Computers & Operations Research, 34(2), pp.578-594. 2007.
- Blom, M., S.O. Krumke, W. de Paepe and L. Stougie. The online-TSP against fair adversaries. In Proc. 4th CIAC 2000, Lecture Notes in Computer Science, Vol. 1767, Springer, Berlin, pp.137-149, 2000.
- Bräysy, O. A Reactive Variable Neighborhood Search for the Vehicle-Routing Problem with Time Windows, Journal on Computing, v15, n4, pp.347-368. 2003.
- Bräysy, O. and W. Dullaert. A fast evolutionary metaheuristic for the vehicle routing problem with time windows, International Journal on Artificial Intelligence Tools, v12, n2, pp.153-172. 2003.
- Bruce, H. Mastering Supply Chain Execution. In APICS International Conference and Exposition, 1999, Ernest N. Morial Convention Center, New Orleans, LA, USA.
- Bruns, A., A. Klose and P. Stähly. Restructuring of Swiss parcel delivery services, OR Spektrum, 22, pp.285-302. 2000.

- Bullo, F., M. Pavone, K. Savla and S.L. Smith. Dynamic Vehicle Routing for Robotic Systems. Proceedings of the IEEE, v99, n9. September 2011.
- Chan, Y., W.B. Carter and M.D. Burnes. A multiple-depot, multiple-vehicle, location-routing problem with stochastically processed demands, Computers and Operations Research, v28, n8, pp.803-826. 2001.
- Chaovalitwongse, W., D. Kim, and P.M. Pardalos. GRASP with a New Local Search Scheme for Vehicle Routing Problems with Time Windows, Journal of Combinatorial Optimization, v 7, n 2, pp.179-207. 2003.
- Chen, F., A. Federgruen and Y.S. Zheng. Coordination mechanisms for a distribution system with one supplier and multiple retailers, Management Science, 47, pp.693–708. 2001.
- Chen, R. Analytical methods for performance enhancement in unreliable multistage manufacturing system with imperfect production. Ph.D Thesis, National University of Singapore. 2010.
- Cheung, R.K. and D.D. Hang. Multi-attribute label matching algorithms for vehicle routing problems with time windows and backhauls, IIE Transactions, 35, pp.191-205. 2003.
- Christofides, N and S. Eilon. Expected distances in distribution problems, Operational Research Quarterly 20, pp437–443. 1969.
- Clarke, G., and J.W. Wright. Scheduling of vehicles from a central depot to a number of delivery points, Operations Research 12, pp.568-581. 1964.

- Confessore, G., D. Corini and G. Stecca. A computational method for pricing of delivery service in a logistics network, International Journal of Production Research, Vol. 46, No. 5, pp.1231–1242. 2008.
- Coslovich, L, R. Pesenti and W. Ukovich. A two-phase insertion technique of unexpected customers for a dynamic dial-a-ride problem, European Journal of Operational Research, 175, pp.1605-1615. 2006.
- Crainic, T.G. and G. Laporte. Planning models for freight transportation, European Journal of Operational Research, 97(3), pp.409–439. 1997.
- Daganzo, C. F and G. F. Newell. Physical distribution from a warehouse: vehicle coverage and inventory levels, Transportation Research Part B: Methodological, v19B, n5, pp.397-408. 1985.
- de Oliveira, S.M., S.R. de Souza and M.A.L. Silva. A Solution of Dynamic Vehicle Routing Problem with Time Window via Ant Colony System Metaheuristic. In 10th Brazilian Symposium on Neural Networks, October 26-30, 2008, Salvador, Bahia, Brazil, pp.21-26.
- Dell'Amico, M., G. Righini, M. Salani. A branch-and-price approach to the vehicle routing problem with simultaneous distribution and collection, Transportation Science, v.40, n.2, pp.235-247. 2006.
- Dong, L., P. Kouvelis and Z. Tian. Dynamic Pricing and Inventory Control of Substitue Products, Manufacturing & Service Operations Management, v11, n2, pp.317-339. 2009.

- Du, L., A. Sharma and S. Peeta. Optimal advance detector location for green termination systems on high-speed isolated rural intersections, Transportation Research Part B: Methodological, v 46, n 10, pp.1404-1418. December 2012.
- Du, T. C., E.Y. Li and D. Chou. Dynamic vehicle routing for online B2C delivery, Omega, v33, i1, pp.33-45. 2005.
- Ebben, M.J.R., E.W. Hans and F.M. Olde Weghuis. Workload based order acceptance in job shop environments, OR Spectrum, 27, pp.107-122. 2005.
- El-Yaniv, R., A. Fiat, R.M. Karp and G. Turpin. Competitive analysis of financial games. In Proc. 33rd Annual Symposium on Foundations of Computer Science, October 1992, Pittsburgh, PA, USA, pp.327-333.
- Ferrucci, F., S. Bock and M. Gendreau. A pro-active real-time control approach for dynamic vehicle routing problems dealing with the delivery of urgent goods. European Journal of Operational Research, V225, I1, pp. 130-141. 2013.
- Fisher, M.L. Optimal Solution of Vehicle Routing Problems Using Minimum K-Trees, Operations Research, v42, n4, pp.626-642. 1994.
- Fleischmann, B., S. Gnutzmann and E. Sandvoβ. Dynamic vehicle routing based on online traffic information, Transportation Science, Vol. 38, No. 4, pp.420-433. 2004.
- Ganesh, K. and T.T. Narendran. CLOVES: A cluster-and-search heuristic to solve the vehicle routing problem with delivery and pick-up, European Journal of Operational Research, v178, i3, pp.699-717. 2007.

- Geunes, J., Z-J. M. Shen and A. Emir. Planning and Approximation Models for Delivery Route Based Services with Price-Sensitive Demands, European Journal of Operational Research, 183, pp.460-471. 2007.
- Ghiani, G., F. Guerriero, G. Laporte and R. Musmanno. Real-time vehicle routing: Solution concepts, algorithms and parallel computing strategies, European Journal of Operational Research, 151(1), pp.1-11. 2003.
- Ghiani, G., A. Quaranta, and C. Triki. New policies for the dynamic traveling salesman problem, Optimization Methods and Software, Vol. 22, No. 6, pp. 971–983. 2007.
- Gillett, S. and L. Miller. A heuristic algorithm for the vehicle dispatch problem, Operations Research, Vol. 22, pp.340-349. 1974.
- Golden, B.L., S. Raghavan and E.A. Wasil, (eds.). The Vehicle Routing Problem: Latest Advances and New Challenges. Operations Research, Computer Science Interfaces Series, v43, Springer, New York, 2008.
- Gribkovskaia, I., Ø. Halskau, G. Laporte and M. Vlček. General solutions to the single vehicle routing problem with pickups and deliveries, European Journal of Operational Research, 180(2), pp.568-584. 2007.
- Gross, D. and C.M. Harris. Fundamentals of Queueing Theory. Thrid edition, New York: John Wiley & Sons, Inc. 1998.
- Haghani, A. and S. Jung. A dynamic vehicle routing problem with time-dependent travel times, Computers & Operations Research, 32, pp.2959–2986. 2005.

- Haimovich, M. and A.H.G. Rinnooy Kan. Bounds and heuristics for capacitated routing problems, Mathematics of Operations Research, 10(4), pp.527–542. 1985.
- Halachmi, B., and W.R. Franta. A Diffusion Approximation to the Multi-Server Queue, Management Science, v24, n5, pp.522-529. 1978.
- Hanshar, F.T. and B.M. Ombuki-Berman. Dynamic Vehicle Routing Using Genetic Algorithms, Applied Intelligence, 27(1), pp.89-99. 2007.
- Hill, A.V., J.M. Hays, E. Naveh. A model for optimal delivery time guarantees, Journal of Service Research, 2(3), pp.254-264. 2002.
- Ho, S.C. and D. Haugland. A tabu search heuristic for the vehicle routing problem with time windows and split deliveries, Computers & Operations Research, v31, n12, pp.1947-1964. 2004.
- Hwang, H.S. Design of supply chain logistics system considering service level, Computers & Industrial Engineering, v43, n1-2, pp.283-297. 2002.
- Jun, Q., J. Wang and B. Zheng. A Hybrid Multi-objective Algorithm for Dynamic Vehicle Routing Problems, In Computational Science - ICCS 2008, 8th International Conference, Part III. Volume 5103 of Lecture Notes in Computer Science, June 2008, Kraków, Poland, June 23-25, 2008. pp.674-681.
- Kimura, T. Diffusion Approximation for an M/G/m Queue, Operations Research, v31, n2, pp.304-321. 1983.
- Kleinrock, L. Queueing Systems, Volume II: Computer Applications, v11, John Wiley & Sons, New York, 1976.
- Kobayashi, H., B. L. Mark, and W. Turin. Probability, Random Processes, and Statistical Analysis. Cambridge; New York: Cambridge University Press, 2012.

- Krumke, S.O., L. Laura, M. Lipmann, A. Marchetti-Spaccamela, W.E. de Paepe, D. Poensgen and L. Stougie. Non-abusiveness helps: an O(1)-competitive algorithm for minimizing the maximum flow time in the online traveling salesman problem. In Approximation Algorithms for Combinatorial Optimization. 5th International Workshop, APPROX 2002, (Lecture Notes in Computer Science Vol.2462), September 2002, Rome, Italy. pp.200-214.
- Kuhn, H.W. The Hungarian Method for the Assignment problem, Naval Research Logistics Quarterly, I2, pp.83-97. 1955.
- Kuo R.J., C.Y. Chiu and Y.J. Lin. Integration of fuzzy theory and ant algorithm for vehicle routing problem with time window. In Annual Meeting of the North American Fuzzy Information Processing Society: Fuzzy Sets in the Heart of the Canadian Rockies (NAFIPS 2004), Jun 27-30 2004, Banff, Alta, Canada. pp.925-930.
- Langevin, A. and P. Mbaraga. Continuous Approximation Models in Freight Distribution: An Overview, Transportation Research Part B: Methodological, v30, n3 pp.163-188. 1996.
- Laporte, G. The Vehicle Routing Problem: An Overview of Exact and Approximate Algorithms, European Journal of Operational Research, 59, pp.345-358. 1992.
- Larsen, A. The Dynamic Vehicle Routing Problem. Ph.D thesis, Department of Mathematical Modelling (IMM), Technical, University of Denmark. 2000.
- Larson, R. and A. Odoni. Urban Operations Research, Englewood Cliffs, New Jersey: Prentice Hall, 1981.

- Lau, H.C., M. Sim and K.M. Teo. Vehicle routing problem with time windows and a limited number of vehicles, European Journal of Operational Research, 148, pp.559-569. 2003.
- Lawler, E.L., J.K. Lenstra, A.H.G. Rinnooy Kan, and D.B. Shmoys (eds.). The Traveling Salesman Problem: A Guided Trip of Combinational Optimization. Chichester, UK: John Wiley, 1985.
- Lehmusvaara, A. Transport time policy and service level as components in logistics strategy: A case study. International Journal of Production Economics, v 56-57, pp.379-387. 1998.
- Lei, L., Q. Wang and C. Fan. Optimal business policies for a supplier-transporter-buyer channel with a price-sensitive demand, Journal of the Operational Research Society, 57, pp281-289. 2006.
- Li, C., D. Simchi-Levi, and M. Desrochers. On the Distance Constrained Vehicle Routing Problem, Operations Research, 40, pp.790-799. 1992.
- Li, J-Q., P.B. Mirchandani, D. Borenstein. Real-time vehicle rerouting problems with time windows, European Journal of Operational Research 194, pp.711–727. 2009.
- Li, Y. and Z. Liu. Study on Decision-Making Mode of Supply Chain with Logistics Time-Sensitive and Price-Sensitive Demand. In Logistics: The Emerging Frontiers of Transportation and Development in China: Proceedings of the Eighth International Conference, October 8-10, 2008, Chengdu, China, pp.1571-1577.

- Lim, A. and F. Wang. Multi-Depot Vehicle Routing Problem: A One-Stage Approach, IEEE Transactions on Automation Science and Engineering, v 2, n 4, pp.397-402. 2005.
- Lysgaard, J., A.N. Letchford and R.W. Eglese. A new branch-and-cut algorithm for the capacitated vehicle routing problem, Mathematical Programming A, 100, pp.423-445. 2004.
- Manasse, M., L.A. McGeoch and D. Sleator. Competitive algorithms for server problems, Journal of Algorithms, 11, pp.208-230. 1990.
- Manski, C.F. and D.L. McFadden. Alternative estimations and sample designs for discrete choice analysis. In Structural Analysis of Discrete Data and Econometric Applications, ed. by C.F. Manski and D.L. McFadden, Cambridge, MA: The MIT Press, 1981.
- Mentzer, J.T., M.B. Myers and M. Cheung. Global market segmentation for logistics services, Industrial Marketing Management, v33, i1, pp.15-20. 2004.
- Mester,D. and O. Bräysy. Active guided evolution strategies for large-scale vehicle routing problems with time windows, Computers & Operations Research, 32, pp.1593-1614. 2005.
- Mitra S. An algorithm for the generalized vehicle routing problem with backhauling, Asia-Pacific Journal of Operational Research, v22, n2, pp.153-169. 2005.
- Monahan, G.E., N.C. Petruzzi and W. Zhao. The Dynamic Pricing Problem from a Newsvendor’s Perspective, Manufacturing & Service Operations Management, Vol.6, No.1, pp.73-91. 2004.

- Montane, F.A.T., R.D. Galvao. A tabu search algorithm for the vehicle routing problem with simultaneous pick-up and delivery service, Computers and Operations Research, v.33, n.3, pp.595-619. 2006.
- Nagarajan, V. and R. Ravi. Approximation Algorithms for Distance Constrained Vehicle Routing Problems, Networks, v59, i2, pp.209-214. 2012.
- Newell, G. F. and C. F. Daganzo. Design of multiple-vehicle delivery trips-l: a ring-radial network, Transportation Research Part B: Methodological, v20B, n5, pp.345-363. 1986.
- Newell, G.F. Approximate Stochastic Behavior of n-Server Service Systems with Large n. Lecture Notes in Economics and Mathematical Systems, n87, Verlag, Berlin: Springer. 1973.
- Pankratz, G. Dynamic vehicle routing by means of a genetic algorithm, International Journal of Physical Distribution & Logistics Management, v.35, n.5, pp.362-83. 2005.
- Papastavrou, J.D. A Stochastic and Dynamic Routing Policy using Branching Processes with State Dependent Immigration, European Journal of Operational Research, 95, pp.167-177. 1996.
- Patron, R.S.F., R.M. Botez and D. Labour. Vertical profile optimization for the Flight Management System CMA-9000 using the golden section search method, IECON 2012 - 38th Annual Conference of IEEE Industrial Electronics (IECON2012), pp.5482-5488. 2012.

- Pavone, M., A. Arsie, E. Frazzoli and F. Bullo. Distributed algorithms for environment partitioning in mobile robotic networks, IEEE Transactions on Automatic Control, v56, n8, pp. 1834-1848. 2011.
- Polacek, M., R.F. Hartl, K. Doerner and M. Reimann. A Variable Neighborhood Search for the Multi Depot Vehicle Routing Problem with Time Windows, Journal of Heuristics, v 10, n 6, pp.613-627. 2004.
- Potvin, J-Y., Y. Xu, I. Benyahia. Vehicle routing and scheduling with dynamic travel times, Computers & Operations Research, 33, pp.1129-1137. 2006.
- Press, W.H., S.A. Teukolsky, W.T. Vetterling, B.P. Flannery. "Section 10.2. Golden Section Search in One Dimension", Numerical Recipes: The Art of Scientific Computing (3rd ed.), New York: Cambridge University Press. 2007.
- Psaraftis, H.N. Dynamic Vehicle Routing Problems. In Vehicle Routing: Methods and Studies, ed. by B.L. Golden and A.A. Assad, pp.223-248. New York: Elsevier Science Publishers B.V. (North Holland), 1988.
- Psaraftis, H.N. Dynamic vehicle routing: Status and prospects, Annals of Operations Research, 61, pp.143-164. 1995.
- Rajeshkumar, B.C. and T. RameshBabu. Evaluation of Logistics Related Policies between Two Different Levels of the Supply Chain Network – a Case Study, Annals of Operations Research, v143, n1, pp.77-89. 2006.
- Renaud, J., G. Laporte and F.F. Boctor. A Tabu Search Heuristic for the Multi-Depot Vehicle Routing Problem, Computers & Operations Research, v23, n3, pp.229-235. 1996.

- Reynolds, J. Logistics and fulfillment for e-business: A Practical Guide to Mastering Back Office Functions for Online Commerce. CMP Books, New York: CMP Media Inc., 2001.
- Ropke, S. Heuristic and exact algorithms for vehicle routing problems. Ph.D. Thesis, Computer science department at the University of Copenhagen (DIKU). December 2005.
- Rosalyn W. CSCMP's 22nd Annual State of Logistics Report, National Press Club, Washington, DC. June 15th, 2011.
- Secomandi, N. A rollout policy for the vehicle routing problem with stochastic demands, Operations Research, v49, n5, pp.796-802. 2001.
- Shinn, S.W. and H. Hwang. Optimal Pricing and Ordering Policies for Retailers under Order Size-Dependent Delay in Payments, Computers and Operations Research, 30, pp.35-50. 2003.
- Sleator, D. and R. Tarjan. Amortized efficiency of list update and paging algorithms, Communications of the ACM, 28(2), pp. 202-208. 1985.
- Steele, J.M. Subadditive Euclidean Functionals and Nonlinear Growth in Geometric Probability, Annals of Probability, 9, pp.365-376. 1981.
- Sunaga, T., E. Kondo and S.K. Biswas. An Approximation Method Using Continuous Models for Queueing Problems, Journal of the Operations Research Society of Japan, v21, pp.29-44. 1978.
- Sural, H., J.H. Bookbinder. The single-vehicle routing problem with unrestricted backhauls, Networks, v.41, n.3, pp.127-136. 2003.

- Taillard, E., P. Badeau, M. Gendreau, F. Guertin and J-Y. Potvin. Tabu search heuristic for the vehicle routing problem with soft time windows, Transportation Science, v31, n2, pp.170-186. 1997.
- Tavakkoli-Moghaddam, R., A.R. Saremi and M.S. Ziaee. A memetic algorithm for a vehicle routing problem with backhauls, Applied Mathematics and Computation, v181, i1, pp.1049-1060. 2006.
- Teodorovic, D. and G. Radivojevic. A fuzzy logic approach to dynamic dial-a-ride problem, Fuzzy Sets and Systems, 116, pp.23-33. 2000.
- Tighe, A., F.S. Smith and G. Lyons. Priority based solver for a real-time dynamic vehicle routing. In 2004 IEEE International Conference on Systems, Man and Cybernetics, vol.7, October 2004, The Hague, Netherlands, pp.6237-6242.
- Tirado, G., L.M. Hvattum, K. Fagerholt, J-F. Cordeau. Heuristics for dynamic and stochastic routing in industrial shipping. Computers and Operations Research, 40, pp.253-263. 2013.
- Viswanathan, S. and Q. Wang. Discount Pricing Decisions in Distribution Channels with Price-Sensitive Demand, European Journal of Operational Research, 149, pp.571-587. 2003.
- Wade, A.C., S. Salhi. An investigation into a new class of vehicle routing problem with backhauls, Omega, v.30, n.6, pp.479-487. 2002.
- Wei, C.T. The Multi-Depot Vehicle Routing Problems, B.Eng Thesis, Department of Mechanical Engineering, National University of Singapore, 2003.

- Wasner, M. and G. Zäpfel. An integrated multi-depot hub-location vehicle routing model for network planning of parcel service, International Journal of Production Economics, v90, i3, pp.403-419. 2004.
- Wieberneit, N. Service network design for freight transportation: a review, OR Spectrum, v30, n1, pp.77-112. 2008.
- Xiang, Z., C. Chu and H. Chen. The study of a dynamic dial-a-ride problem under time-dependent and stochastic environments, European Journal of Operational Research, 185, pp.534-551. 2008.
- You, P. Ordering and pricing of service products in an advance sales system with price-dependent demand, European Journal of Operational Research, 170, pp.57-71. 2006.
- Zhong Y., and M.H. Cole. A vehicle routing problem with backhauls and time windows: a guided local search solution, Transportation Research Part E: Logistics and Transportation Review, v.41, i.2, pp.131-144. 2005.
- Zhou, W. and C. Lee. Pricing and competition in a transportation market with empty equipment repositioning, Transportation Research Part B, 43, pp.677–691. 2009.
- Zhu, K.Q. and K.L. Ong. A reactive method for real time dynamic vehicle routing problem. In Proceedings 12th IEEE Internationals Conference on Tools with Aritficial Intelligence, ICTAI 2000, November 2000, Vancouver, BC, Canada, pp176-180.

- Zuo, Q., X. Yin, J. Zhou, B-J. Kwak and H.K. Chung. Implementation of Golden Section Search Method in SAGE Algorithm, 2011 5th European Conference on Antennas and Propagation (EuCAP), pp.2028-2032. 2011.
- Logistics and Supply Chain Management (SCM) Key Performance Indicators (KPI) Analysis: A Canada / United States Manufacturing Perspective, November, 2006.
- Repowering Transport: Project White Paper, World Economic Forum, April, 2011.

Appendix A. Construction of the Intensity Matrix

The balance equations can be summarized as the stationary Equation (3.15). In the stationary equation, the intensity matrix indicates the transition rate between all system states. The elements in the intensity matrix are listed in the following sections for different cases discussed in chapters 3 and 4.

A.1. Intensity Matrix in the Analysis of the Vehicle Departure Issue

The construction of the intensity matrix for the customer waiting process in the analysis of the vehicle departure issue is summarized as follows.

The transition rate from (*w*,*k*,*I*,*b*) to (*w*,*k*,*I*-1,*b*) is μ_{I-1}, when the vehicle is travelling between customers.

1st trip: $q'_{(w,k,I,b),(w,k,I-1,b)} = \mu_{I-1}$, $w = 1, 2, \cdots, N_D; k = 1, 2, \cdots; I = 1, 2, \cdots; b = 1, 2, \cdots, N_D$

2nd trip: $q'_{(0,k,I,0),(0,k,I-1,0)} = \mu_{I-1}$, $k = 1, 2, \cdots; I = 1, 2, \cdots$

When the vehicle is travelling between the warehouse and a customer on the first trip, the transition rate is $\mu'_{I-1}/2$. It is μ'_{I-1} for the second trip.

$$q'_{(w,0,I,b),(w,0,I-1,b)} = \mu'_{I-1}/2,\quad w = 1, 2, \cdots; I = 2, 3, \cdots; b = 1, 2, \cdots, w$$

$$q'_{(0,0,I,0),(0,0,I-1,0)} = \mu'_{I-1},\quad I = 1, 2, \cdots$$

When there are at least N_D customers in the waiting list, the vehicle can start the second trip. The transition rate is $\mu'_0/2$.

$$q'_{(N_D,0,1,b),(0,b,0,0)} = \mu'_0/2,\quad b = 1, 2, \cdots, N_D$$

$$q'_{(w,0,1,w),(0,w,0,0)} = \mu'_0/2,\quad w = N_D + 1, N_D + 2, \cdots$$

If there are not enough customers in the waiting list to trigger the next trip, the vehicle has to stay at the warehouse.

$$q'_{(w,0,1,b),(w,0,0,b)} = \mu'_0/2, \quad w = 1,2,\cdots,N_D - 1; b = 1,2,\cdots,w$$

The system transitions from state (w,k,0,b) to (w,k-1,I_1,b) with a transition rate μ_{I_1}, when the vehicle finishes the service for a customer and sets out to the next destination.

$$q'_{(w,k,0,b),(w,k-1,I_1,b)} = \mu_{I_1}, \quad w = 0,1,\cdots,N_D; k = 2,3,\cdots; b = 0,1,\cdots,w$$

The system transitions from state (w,1,0,b) to (w,0,I_0,b) with a transition rate $\mu'_{I_0}/2$, when the vehicle finishes the service for the last customer on the first trip and starts heading to the warehouse.

$$q'_{(w,1,0,b),(w,0,I_0,b)} = \mu'_{I_0}/2, \quad w = 1,2,\cdots,N_D; b = 1,2,\cdots,w$$

The transition rate is μ'_{I_0}, when the vehicle starts the last leg of the second trip.

$$q'_{(0,1,0,0),(0,0,I_0,0)} = \mu'_{I_0}$$

When the vehicle finishes the service for the specific customer, the process terminates at state (0,0,0).

$$q'_{(0,0,1,0),(0,0,0,0)} = \mu'_0$$

If $0 < w < N_D$, the system still counts for new demands until $w = N_D$.

$$q'_{(w,k,I,b),(w+1,k,I,b)} = \lambda, \quad w = 1,2,\cdots,N_D - 1; k = 0,1,\cdots; I = 0,1,\cdots; b = 1,2,\cdots,w$$

The diagonal elements of the intensity matrix are calculated as follows.

$$q'_{(w,k,I,b),(w,k,I,b)} = -\mu_{I-1} - \lambda, \quad w = 1,2,\cdots N_D - 1; k = 1,2,\cdots; I = 1,2,\cdots; b = 1,2,\cdots N_D - 1$$

$$q'_{(N_D,k,I,b),(N_D,k,I,b)} = -\mu_{I-1}, \quad k = 1,2,\cdots; I = 1,2,\cdots; b = 1,2,\cdots N_D - 1$$

$$q'_{(0,k,I,0),(0,k,I,0)} = -\mu_{I-1}, \quad k = 1,2,\cdots; I = 1,2,\cdots$$

$$q'_{(w,k,0,b),(w,k,0,b)} = -\mu_{I_1} - \lambda, \quad w = 1,2,\cdots N_D - 1; k = 2,3,\cdots; b = 1,2,\cdots N_D - 1$$

$$q'_{(N_D,k,0,b),(N_D,k,0,b)} = -\mu_{I_1}, \quad k = 2,3,\cdots; b = 1,2,\cdots N_D - 1$$

$$q'_{(w,1,0,b),(w,1,0,b)} = -\lambda - \mu'_{I_0}/2, \quad w = 1,2,\cdots N_D - 1; b = 1,2,\cdots N_D - 1$$

$$q'_{(N_D,1,0,b),(N_D,1,0,b)} = -\mu'_{I_0}/2, \quad b = 1,2,\cdots N_D - 1$$

$$q'_{(0,k,0,0),(0,k,0,0)} = -\mu_{I_1}, \quad k = 2,3,\cdots$$

$$q'_{(0,1,0,0),(0,1,0,0)} = -\mu'_{I_0}$$

$$q'_{(w,0,I,b),(w,0,I,b)} = -\lambda - \mu'_{I-1}/2, \quad w = 1,2,\cdots N_D - 1; I = 1,2,\cdots; b = 1,2,\cdots, w$$

$$q'_{(N_D,0,I,b),(N_D,0,I,b)} = -\mu'_{I-1}/2, \quad I = 1,2,\cdots; b = 1,2,\cdots, N_D$$

$$q'_{(w,0,I,w),(w,0,I,w)} = -\mu'_{I-1}/2, \quad w = N_D + 1, N_D + 2,\cdots; I = 1,2,\cdots$$

$$q'_{(0,0,I,0),(0,0,I,0)} = -\mu'_{I-1}, \quad I = 1,2,\cdots$$

$$q'_{(0,0,0,0),(0,0,0,0)} = 0$$

A.2. Intensity Matrix in the Analysis of the Vehicle Capacity Issue

In the analysis of the vehicle capacity issue, the construction of the intensity matrix for the customer waiting process is summarized as follows.

The transition rate from (*w*,*k*,*I*,*b*) to (*w*,*k*,*I*-1,*b*) is μ_{I-1}, when the vehicle is travelling between customers.

$$q'_{(w,k,I,b),(w,k,I-1,b)} = \mu_{I-1}, \quad w = 1,2,\cdots; k = 1,2,\cdots; I = 1,2,\cdots; b = 1,2,\cdots, C$$

$$q'_{(0,k,I,0),(0,k,I-1,0)} = \mu_{I-1}, \quad k = 1,2,\cdots; I = 1,2,\cdots$$

When the vehicle is travelling between the warehouse and a customer on the trip for the specific customer, the flow rate is μ'_{I-1}. However, it is $\mu'_{I-1}/2$ for previous trips.

$$q'_{(0,0,I,0),(0,0,I-1,0)} = \mu'_{I-1}, \quad I = 1,2,\cdots$$

$$q'_{(w,0,I,b),(w,0,I-1,b)} = \mu'_{I-1}/2, \quad w = 1,2,\cdots; I = 2,3,\cdots; b = 1,2,\cdots,C$$

When there are more than C customers in the waiting list at the time the vehicle starts a new trip, the trip for the specific customer has not been scheduled yet.

$$q'_{(w,0,1,b),(w-C,C,0,b)} = \mu'_0/2, \quad w = C+1, C+2,\cdots; b = 1,2,\cdots,C$$

When there are $N_D \leq w \leq C$ customers in the waiting list, the vehicle starts a trip for the specific customer. The flow rate is $\mu'_0/2$.

$$q'_{(N_D,0,1,b),(0,b,0,0)} = \mu'_0/2, \quad b = 1,2,\cdots,N_D$$

$$q'_{(w,0,1,w),(0,w,0,0)} = \mu'_0/2, \quad w = N_D+1, N_D+2,\cdots,C$$

If there are not enough customers in the waiting list to trigger the trip for the specific customer, the vehicle has to stay at the warehouse.

$$q'_{(w,0,1,b),(w,0,0,b)} = \mu'_0/2, \quad w = 1,2,\cdots,N_D-1; b = 1,2,\cdots,C$$

The system transitions from state (w,k,0,b) to (w,k-1,I_1,b) with a flow rate μ_{I_1}, when the vehicle finishes the service for a customer and sets out to the next destination.

$$q'_{(w,k,0,b),(w,k-1,I_1,b)} = \mu_{I_1}, \quad w = 0,1,\cdots,N_D; k = 2,3,\cdots; b = 0,1,\cdots,C$$

The system transitions from state (w,1,0,b) to (w,0,I_0,b) with a flow rate $\mu'_{I_0}/2$, when the vehicle leaves the last customer on the current trip and starts heading to the warehouse.

$$q'_{(w,1,0,b),(w,0,I_0,b)} = \mu'_{I_0}/2, \quad w = 1,2,\cdots,N_D; b = 1,2,\cdots,C$$

The flow rate is μ'_{I_0}, when the vehicle is on the last leg of the trip between the warehouse and a customer.

$$q'_{(0,1,0,0),(0,0,I_0,0)} = \mu'_{I_0}$$

When the vehicle finishes the service on the last trip, the process terminates at state (0,0,0).

$$q'_{(0,0,1,0),(0,0,0,0)} = \mu'_0$$

If the remainder of w-1 divided by C is less than N_D-1, the system will continue counting new demands until w reaches the nearest N_D+Cj.

$$q'_{(w,k,I,b),(w+1,k,I,b)} = \lambda, \quad w = Cj+1, Cj+2, \cdots, Cj+N_D-1; b = 1,2,\cdots, N_D-1; k, I, j = 0,1,\cdots$$

The diagonal elements of the intensity matrix are calculated as follows.

$$q'_{(w,k,I,b),(w,k,I,b)} = -\sum_{v \neq (w,k,I,b)} q'_{(w,k,I,b),v}, \ \forall (w,k,I,b)$$

A.3. Intensity Matrix in the Analysis of Routing Strategies

A.3.1. Intensity Matrix for the Steady State Process

At any moment, the state may transition from (w,I) to (w,I-1) with a flow rate μ_{I-1}, when the vehicle travels along the road.

$$q_{(w,I),(w,I-1)} = \mu_{I-1}, \quad w = 0,1,\cdots; I = 1,2,\cdots, n$$

Due to the redundant state (w,0), the state may transition directly from (w,1) to (0,I) with a flow rate $\mu_0 p_{w,I}$, when the vehicle returns to the warehouse and starts the next trip.

$$q_{(w,1),(0,I)} = \mu_0 p_{w,I}, \quad w = 1,2,\cdots$$

The system stays in the state (0,0), if there is no customer waiting for the service.

$$q_{(0,1),(0,0)} = \mu_0$$

When a new customer appears, the number of customers in the system increases by one.

$$q_{(w,I),(w+1,I)} = \lambda, \quad w = 0,1,\cdots; I = 0,1,\cdots \text{ except } w = I = 0$$

If a new customer appears when the vehicle is idle, the vehicle will immediately set out to the location of this new customer. The state transitions from (0,0) to (0,I) with a flow rate $\lambda p_{1,I}$.

$$q_{(0,0),(0,I)} = \lambda p_{1,I}, \quad I = 1,2,\cdots,n$$

The diagonal elements of the intensity matrix are calculated as follows.

$$q_{(w,I),(w,I)} = -\sum_{v \neq (w,I)} q_{(w,I),v}, \quad \forall (w,I)$$

A.3.2. Intensity Matrix for the Customer Waiting Process

The transition rate for the vehicle travelling along the road is μ_I on the first trip, $\mu_{w,I}$ on the second trip between customers, and $\mu'_{w,I}$ on the second trip between the warehouse and customers.

$$q'_{(w,-1,I+1),(w,-1,I)} = \mu_I, \quad w = 1,2,\cdots; I = 1,2,\cdots,n$$

$$q'_{(w,k,I+1),(w,k,I)} = \mu_{w,I}, \quad w = 1,2,\cdots; k = 1,2,\cdots,w; I = 0,1,\cdots,I_{1,w} - 1$$

$$q'_{(w,0,I+1),(w,0,I)} = \mu'_{w,I}, \quad w = 1,2,\cdots; I = 1,2,\cdots I_{0,w} - 1$$

The specific customer can be scheduled on any leg of the second trip with an equal chance.

$$q'_{(w,-1,\ ),(w,k,0)} = \mu_0 / w, \quad w = 1,2,\cdots; k = 1,2,\cdots,w$$

On the second trip, the system transitions from state $(w,k,0)$ to $(w,k\text{-}1,I_{1,w})$ with a flow rate $\mu_{w,I_{1,w}}$, when the vehicle finishes the service of the current customer and sets out to the next destination.

$$q'_{(w,k,0),(w,k-1,I_{1,w})} = \mu_{w,I_{1,w}}, \quad w = 1,2,\cdots; k = 2,3,\cdots,w$$

The system transitions from state $(w,1,0)$ to $(w,0,I_{0,w})$ with a flow rate $\mu'_{w,I_{0,w}}$

$$q'_{(w,1,0),(w,0,I_{0,w})} = \mu'_{w,I_{0,w}}, \quad w = 1,2,\cdots$$

When the vehicle finishes the service for the specific customer, the process terminates at state (0,0,0).

$$q'_{(w,0,1),(0,0,0)} = \mu'_{w,0}\,, \quad w = 1, 2, \cdots$$

The diagonal elements of the intensity matrix are calculated as follows.

$$q'_{(w,k,I),(w,k,I)} = -\sum_{v \neq (w,k,I)} q_{(w,k,I),v}\,, \quad \forall (w,k,I)$$