Courier Collaboration in Pickup and Delivery Services by Hub Transshipment across Flexible Time Periods

Yao-Hung Chen and Yon-Chun Chou
National Taiwan University

Abstract

Most research work on pickup and delivery routing problems is concerned with developing solution algorithms. However, when the service area is large and job density is low, couriers frequently travel a long distance to serve a few customers. Service inefficiency can be attributed to inherent lacking of economies of scale and reflects on tough tradeoffs between courier utilization and customer waiting time. In this paper, operational policy design of courier services is addressed aiming at reducing both workload and waiting time. Regression metamodels of tour length are first constructed by simulation. Mean-value performance analysis of a new policy of hub transshipment across flexible time period contrast to the periodical routing policy is next presented. Applicability condition of the new policy is provided. Finally, dynamic operation of the policy is illustrated with styled data of courier service at a large hospital.

Keywords: Pickup and delivery routing, transshipment, flexible time period, collaborative services, multiple objectives.

1. Introduction

Pickup and delivery (P&D) routing problems are widely studied in courier service and transportation logistics. In most problem settings, the constraints of vehicle capacity, tour time, waiting time, time window, and precedence relation are taken into consideration. Routing decisions typically include clustering of customers, customer-vehicle assignment, and P&D sequencing. Some models allow for transshipment, which permits a customer to be picked up by one vehicle and transshipped to another vehicle. The literature is extensive (see Parragh et al. [14] and Berbeglia et al. [3]). Most research work is devoted to developing solution algorithms and meta-heuristics, given that the routing problems are NP-hard. However, when the service area is large and job density is low, couriers frequently travel a long distance to serve a few customers. Inefficiency of service can be attributed to inherent lacking of economies of scale. To address tough tradeoffs between courier utilization and customer waiting time, the approach of operational policy design is adopted in this paper, as compared to the approach of routing optimization.

P&D routing models can be classified as either static or dynamic. In static models, a set of P&D jobs is given as input. In dynamic models, jobs continue to arrive even after the P&D service has commenced. Therefore, static models are used for a single period of finite duration and dynamic models are used for one continuous time period. There has been little study on multi-period P&D routing problems. The problems arise when routing decisions span multiple time periods because jobs are either postponable (see Angelelli et al. [1]) or have extendable deadline (see Wen et al. [17]). In this paper, a courier routing problem of multiple time periods of flexible length with transshipment between time periods is addressed. The problem arises in a courier service problem of a large hospital. We will next describe the service problem before presenting relevant literature survey.

The hospital that we studied is a full-service hospital with 24 clinical departments and more than 200 clinical rooms in sprawling buildings. After visiting a clinic, some patients are referred to other departments for further treatment. Before a patient can be examined by a second physician, however, some medical materials of the patient must be transported to the second clinical department through a courier service. Because of the inefficiency in P&D services, patients sometimes have to wait at the second department for the delivery service. Patient's waiting time is considered a very important quality measure. Because the number of couriers is constrained, the challenge is to improve patient waiting time by designing creative routing solutions without increasing the courier staff. This problem has three characteristics: (1) large service territory, (2) uncertain P&D locations, and (3) limited resource of couriers. Furthermore, both patient waiting time and courier utilization are important performance criteria. We approach this problem by redesigning operational policy, which is a task preceding the development of routing algorithms. The following literature review will focus on salient characteristics and innovative policy design for the P&D routing problems. We use the words jobs, requests and demands interchangeably to refer to customer requests of P&D services.

There are two solution approaches to dynamic P&D routing problems (see Berbeglia et al. [3]). The first approach is to solve a static problem each time a new request arrives. The second approach is to solve a static problem at the beginning to obtain an initial solution and then, with each new request, to revise the solution by using heuristics of inserting and rearranging route segments. Because job arrivals are uncertain, associated with the decision of routing is vehicle prepositioning in anticipation of future arrivals. In the dynamic environment, the most prominent strategy is related to that of waiting. Mitrovic-Minic and Laporte [11] compared three waiting strategies with the drive-first strategy. Their simulation study showed that the three waiting strategies outperform the drive-first strategy on tour length but, in some simulation runs, at the expense of slightly larger fleet. Pureza and Laporte [15] evaluated the effect of a waiting strategy and a request buffering strategy in dynamic P&D problems with time window constraints and uncertain travel time between each pair of locations. Both strategies are postponement strategies. While the waiting strategy is a policy that delays the assignment of vehicles to their next service destination, the buffering strategy is a policy for aggregating non-urgent requests before they are served in a continuous sub-tour. Their simulation results validated the advantage of the two strategies over the traditional drive-first strategy.

Transshipment provides opportunities for multiple vehicles to collaborate. By adding flexibility to routing, it has positive effects on reducing the waiting time of the customers and the travel cost of the vehicles. Nakao and Nagamochi [13] did a worst case analysis of cost savings when a transshipment point is introduced. They showed that the bounds are in proportion to the square root of the number of routes and the square root of the number of requests. Cortés et al. [4] developed a branch-and-cut algorithm based on Benders decomposition for the P&D problem with transshipment. They compared the computational efficiency of their algorithm with a straight branch and bound algorithm. By experimenting with small problem instances, they concluded that there exist some configurations in which transshipment can be more profitable and further conjectured that transshipment would be effective under high demand conditions. Mitrovic-Minic and Laporte [12] applied and evaluated the policy of transshipment on a P&D problem with time window constraints. By using heuristics and simulation with stylized data, they showed that the policy leads to a reduction in the total travel distance when requests are uniformly generated in the plane. The benefit is more significant when the problem size is large and requests are clustered. In another application with stylized data, Lin [8] evaluated the benefits of transshipment in local courier service of a multi-national logistics firm. Each request has a pickup time window and a delivery deadline at the depot. The objective is to minimize the sum of fixed and operation costs. The flexibility of transshipment leads to a cost savings of approximately 10%-20%, depending on operation modes.

Information about future arrivals can be exploited to improve routing decisions (see Liu and Xu [9]). Larsen et al. [7] analyzed the effect of arrival dynamism on the performance of routing heuristics. The dynamism is defined as the ratio of dynamic requests over the total number of requests. They applied the approach of stochastic analysis by constructing functional relationship between performance measures and the degree of dynamism. They found that the nearest neighbor heuristic uniformly outperforms other policies studied.

P&D problems in multiple time periods have received very little attention in the literature. Athanasopoulos and Minis [2] addressed an appointment-based courier service problem in which some service requests must be fulfilled in specified periods while others can be fulfilled within a multi-period horizon. They proposed a method for the assignment of service requests over a rolling horizon. In this paper, we also address a P&D problem of multiple periods. But, the time periods have flexible length. The salient idea is to have multiple couriers collaborate through transshipment across time periods to improve both workload and waiting time. This idea is probable since workload could be reduced with transshipment and reduced workload in turn could lead to an increase in service frequency and a reduction in waiting time. It is the objective of this study to validate this plausibility by a formal analysis.

This paper presents a method for analyzing transshipment collaboration of multiple couriers with flexible time periods. For brevity of terminology, we call the total time that a customer spends in the system the sojourn time, following the convention in queuing theory. In P&D applications that involve human customers, waiting time or sojourn time is usually considered the most important criterion of routing decisions. In other applications, route length or work time is usually considered the primary criterion of operation efficiency. However, sojourn time and workload are not unrelated but are frequently conflicting objectives (see Liu and Xu [9]). A contribution of this paper lies in an explicit treatment of the tradeoffs between sojourn time and workload.

The rest of this paper is organized as follows. In Section 2, a problem of P&D service by multiple couriers over a large territory is defined. The current operation policy and a new policy design are described. In Section 3, a regression metamodel of the tour length is constructed by simulation. In Section 4, analytical results are presented for the applicability of the new policy. In Section 5, dynamic evaluation of the policy is illustrated with styled data of courier service in a large hospital. Finally, conclusions and discussions can be found in Section 6.

2. Problem Description

There are many variants of the P&D problems. The problem addressed in this paper has the following characteristics (or premises):

(1) A large service territory is divided into a number of regions, each served by a courier. All couriers are stationed at a central depot. Couriers depart from the depot for each service tour and must return to the depot at the end of the tour.

(2) Service requests follow a Poisson process of arrival and they are uniformly

distributed in the territory. Each request has an origin and a destination. The origin and destination of a job might be in different regions or the same region. Services are not provided immediately after they are requested. Instead, service requests are batched.

(3) Transshipment takes place at the depot. Jobs to be transshipped are exchanged among couriers at the depot.

(4) Both customer sojourn time and courier workload are important criteria of performance.

This study is motivated by the challenges of dual objectives on courier workload and customer sojourn time. The courier staff is constrained; improving service response time by increasing the courier staff is precluded. The research approach taken by this study is not on developing advanced routing algorithms. Instead, the focus is on re-designing operation policies. Specifically, the expected performance of the following two policies is compared. Policy 1 is very common in practice and is well studied in static routing models. Policy 2 is a new design, but requires detailed analysis of its effect.

(1) Policy 1 (periodical routing): Requests are accumulated and then dispatched to couriers at periodical intervals. The couriers pick up all orders that originate in their assigned regions and do all the deliveries.

(2) Policy 2 (hub transshipment with flexible time periods): Couriers pick up jobs in their duty regions but deliveries are restricted to those in their duty regions. Jobs which are destined for other regions are brought back to the depot and are delivered by another courier in the next period. This policy saves couriers from making lengthy excursions from their duty region.

We choose to use tour length as the measure for efficient resource use. For variable period length, courier utilization will change with the period length and therefore is not an appropriate measure. In contrast, tour length is fundamental. Tour time and work time can be derived from tour length. We will use the terms tour length, tour time and courier workload interchangeably in discussion.

3. Tour Length Function

It is common in stochastic routing research to construct metamodels for the input data of optimization models or for the model output. Winch et al. [18] used a regression metamodel in a goal programming model of reverse logistics. Larsen et al. [7] constructed functional relationships between performance measures and dynamism of job arrivals. In this section, a regression metamodel of tour length as a function of both job quantity and territory size is constructed by simulation.

The main decision variable of policy 2 is the length of time period, which will also be called time bucket size or bucket size. The bucket size determines the quantity of jobs that are accumulated for each service tour. It is well understood that the efficiency of routing is strongly related to the total number or the density of jobs. The higher the average number of jobs served per unit time or unit length of tour, the larger the economies of scale. In this section, the tour length as a function of job quantity and territory size is derived from simulated data and regression. This function is an input to policy design in Section 4. A generic service territory configuration as shown in Figure 1 is used in the analysis. In the configuration, three unit circles represent the duty regions of three couriers. The depot is located at the geometric center of the three circles.

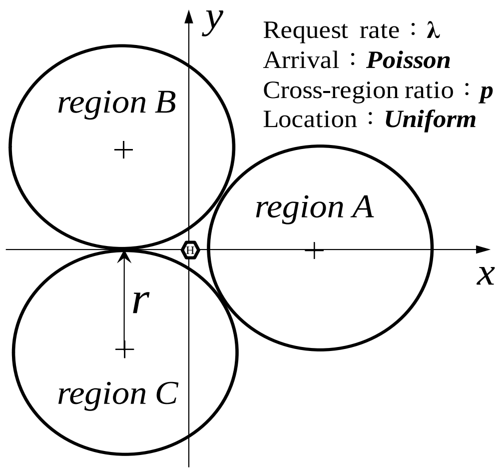

Figure 1: A generic P&D territory.

For N points that are uniformly and independently scattered in a connected region of area A, the expected tour length, D, can be derived as:

$$D \approx \phi \cdot \sqrt{AN}, \text{ as } N \to \infty \tag{3.1}$$

where ϕ is a constant (see Daganzo [5]). In another study Equation (3.1) is also obtained from the optimal tours for random Traveling Salesman Problems with N cities uniformly distributed over a rectangular area (see Gent and Walsh [6]). It is interesting to note that these studies consistently show that the tour length D is approximately proportional to $N^{1/2}$. We will call this relationship between D and N the square root formula.

In this study, a metamodel is constructed for a number of reasons. First, P&D routing problems are more constrained than general routing problems. A job can only be delivered after it has been picked up. The tour length is therefore necessarily longer than without such a constraint. It would be interesting to know if the square root formulas still offer good approximation for P&D problems. The second reason is based on practical considerations. The shape of the territory of the hospital that is studied does not resemble a circle or a rectangle. The locations of jobs are clustered in several unconnected areas, each covering several buildings. We would argue that the P&D service in most cities have similar characteristics of the territory. Most cities are composed of several sub-cities which are separated by parks, rivers and other spaces in which no P&D jobs arise. The third reason is that our problem includes the cross-region probability. Our model explicitly includes two types of jobs: intra-region and cross-region. It would be interesting to know if the square root formula is still valid under the influence of job transshipment.

Using a territory with 3 clusters as in Figure 1 has another contribution. A territory with 2 clusters resembles a rectangle in the outer perimeter. A territory with 4 or more clusters also resembles a rectangle or circle. If we have to pick just one configuration for this study, the configuration of Figure 1 would be distinct from the territory shape that has been typically dealt with in the literature, i.e., rectangular and circular.

The first task of policy comparison is to construct a tour length model for varying batch size (or job quantity). In the literature, operation and control policies are usually studied by using stochastic routing analysis. Job arrival is typically assumed to follow a Poisson process and pickup and destination locations are uniformly distributed in space (see Swihart and Papastavrou [16] and Mes et al. [10]). In this study, the tour length model is constructed by using simulation with the following procedure.

(1) Generate random jobs from the territory configuration of Figure 1. Jobs that are picked up and delivered in the same region are called same-region (*SR*) jobs. Jobs that are delivered out of their pickup regions are called cross-region (*CR*) jobs. The fraction (p) of *CR* jobs is treated as a parameter.
(2) Apply the Nearest Neighbor (NN) heuristic to obtain solution routes. Because P&D routing problems are NP-hard, heuristic methods are used in most applications. Therefore, in this metamodeling task, a heuristic method is used as the solution engine. The NN heuristic is chosen as it has been shown to produce good results, e.g., in the work of Swihart and Papastavrou [16] and Larsen et al. [7] reviewed above.
(3) Steps 1 and 2 are repeated 5 times for each batch size (1 to 40). The total number of data point is 200.
(4) Apply regression analysis on 200 data points to obtain a functional relationship between average tour length and batch size.

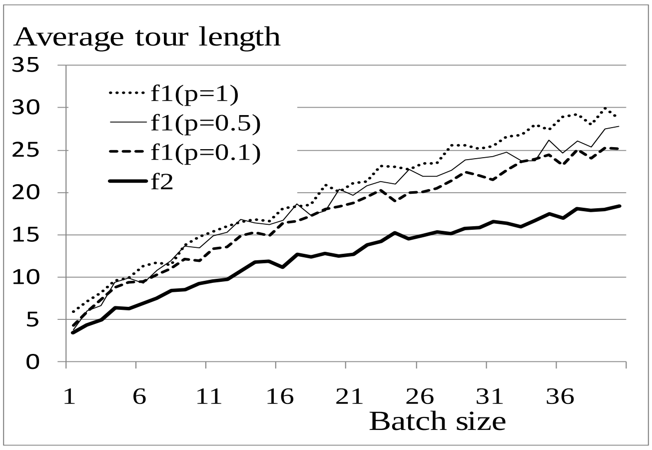

Figure 2: Simulated data for route length.

The regression results are shown in Figure 2. The function f_2 is a route length model for job sets that contain only *SR* jobs. The function f_1 is a route length model for job sets that contain both *SR* and *CR* jobs. For instance, the case of p=1 indicates that all jobs are *CR* jobs and p=0.1 indicates that 10% of the jobs are *CR* jobs. The fitted functions are concave and increasing in batch size. There are economies of scale and their effect is diminishing with batch size. By applying regression analysis, the following power functions, where q is the batch size, are obtained for the four curves.

$$f_1(q) = a_1 q^{c_1} = \begin{cases} 4.2957 * q^{0.5237}, if \ \ p = 1.0 \quad (R^2 = 0.932) \\ 4.0241 * q^{0.5214}, if \ \ p = 0.5 \quad (R^2 = 0.948) \\ 4.0156 * q^{0.5008}, if \ \ p = 0.1 \quad (R^2 = 0.941) \end{cases} \tag{3.2}$$

$$f_2(q) = a_2 q^{c_2} = 3.0903 * q^{0.4797}, \qquad (R^2=0.938) \tag{3.3}$$

For the convenience of analysis, we will treat q as a real-valued variable, $q \geq 0$. The function f_2 is at a level lower than f_1, as f_2 is for a smaller service region than f_1. That is, $f_2(q) \leq f_1(q)$. In addition, $a_2 < a_1$, $c_2 < c_1 < 1$ and $f_1(0) = f_2(0) = 0$.

In stochastic routing analysis, a tour is composed of a number of segments, which are generated from random points. We conjecture that the length of P&D tours is normally distributed. We used a second simulation experiment to validate this conjecture. The minimum number of P&D requests is one, which requires a tour of two nodes (n=2q). The simulation settings are r=1, and p=0.5. The number of experiment replications is set at 30, as the t-test is applied. Denote the number of service requests as q. Figure 3 summarizes the simulation results of tour lengths L_1 and L_2 for q=3, 4 and 5. Both Kolmogorov-Smirnov test and Shapiro-Wilk normality test cannot reject the normality hypothesis at type I error of 0.05 (data in Table 1). Due to some simulation study and the support of the following statistical tests analysis, we would like to give the following

Conjecture 1. *For Poisson arrival of service request, the tour length is normally distributed.*

Supporting evidence. It's supported by the normality test on simulation results (Figure 3 and Table 1). Conjecture 1 is stipulated for up to 10 nodes. This range of jobs is sufficient for the application of hospital courier services of this study. □

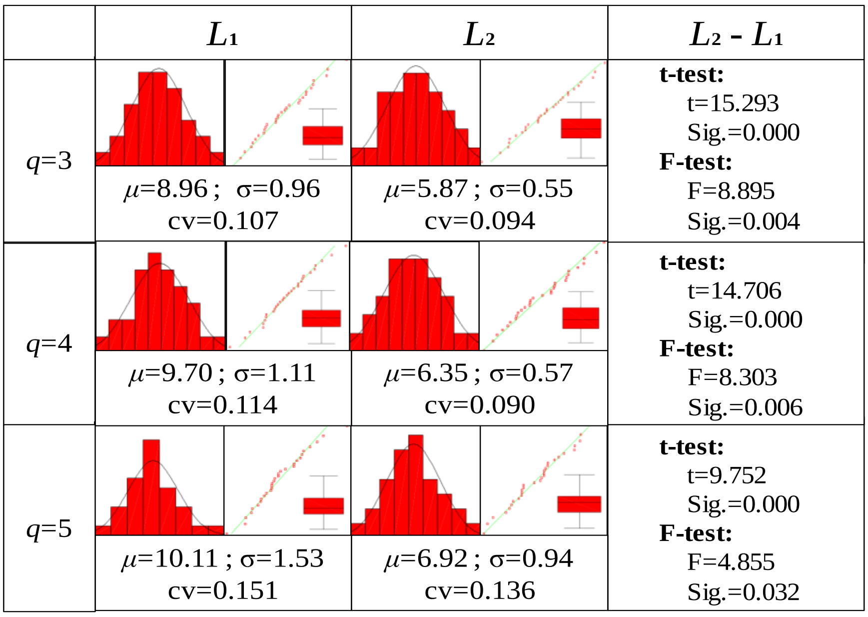

Figure 3: Tour length normality.

Table 1: Resultant data of normality tests.

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
q=3	L_1	.085	30	.200*	.984	30	.925
	L_2	.065	30	.200*	.990	30	.993
q=4	L_1	.081	30	.200*	.994	30	1.000
	L_2	.103	30	.200*	.977	30	.732
q=5	L_1	.115	30	.200*	.979	30	.801
	L_2	.112	30	.200*	.981	30	.840

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction.

The route length functions $f_1(q)$ and $f_2(q)$ are obtained from the territory configuration of three unit circles (Figure 1). They can be generalized to cases with circles of radius r, $r \neq 1$, in the following Lemma 2. Therefore, using unit circles in our analysis is not a limiting assumption. The research results generalize easily to cases of $r \neq 1$.

Lemma 2. (*Scaling effect of service regions*) *If the service regions are not unit circles but are circles of radius r, the tour length functions are* $r \cdot f_1(q)$ *and* $r \cdot f_2(q)$.

Proof. Considering a r times magnification of 3-circle topology configuration in Figure 1. All the unit-circles will become circles of radius r. With result of affine transformation, the length of each tour segment in the configuration is expanded by a factor r, so are the tour length functions. □

4. Performance Analysis

Analytical results for comparing policies 1 and 2 are presented in this section. Both polices take a varying batch of jobs as input. Mean value analysis is commonly used in queuing analysis and manufacturing system design. Although job arrivals are dynamic in this study, performance comparison is based on mean-value analysis for two reasons. First, policy design is not a routine decision. Second, excursions from the mean value can be coped with by dynamic routing or other reactive measures in actual operation. The dynamic execution of policy 2 will be addressed in section 5.

The sojourn time and work time can be estimated by using tour length functions. Besides the radius r, our analysis includes two more parameters: job arrival rate λ and cross-region job probability p. The period length b is a decision variable with a policy subscript i, $i \in \{1, 2\}$. The courier speed v appears in the analysis. But it will become clear that it is just a scaling parameter for relating tour length to tour time. Notations related to the performance measures are:

$L_i(q)$: tour length under policy i, where job quantity q is also expressed as λb_i.

T_i: sojourn time per job under policy i.

For policy 1, the tour length can be obtained directly from the function f_1. The batch size equals to the total number of jobs that arrive during a period:

$$L_1 = f_1(\lambda b_1) \cdot r \tag{3.4}$$

Following policy 2, each courier will give p fraction of jobs to other couriers and receive an equal amount from other couriers as jobs are uniformly distributed in all regions. The route length is given by function f_2:

$$L_2 = f_2(\lambda b_2) \cdot r \tag{3.5}$$

For a time bucket of b, a job will spend an average time of b/2 waiting for the next delivery to commence. Also, on average, a job will stay on a delivery tour for one half of the tour time. For policy 1, the average sojourn time is:

$$T_1(b_1) = \frac{1}{2}b_1 + \frac{1}{2}L_1 / v = \frac{1}{2}b_1 + \frac{1}{2}f_1(\lambda b_1) \cdot r / v \equiv \frac{1}{2}b_1 + \frac{1}{2}f_1(\lambda b_1) \cdot s \tag{3.6}$$

The ratio r/v is the nominal size of the territory. Denote it as s. For policy 2, there are two types of jobs to consider. The fraction of CR jobs is p and that of SR jobs is 1-p. The sojourn time T_2 is a weighted average of the sojourn times of these two types of jobs. The delivery of CR jobs will be delayed by one period. Thus,

$$\begin{aligned} T_2(b_2) &= [\frac{1}{2}b_2 + b_2 + \frac{1}{2} f_2(\lambda b_2)\cdot r/v]\cdot p + [\frac{1}{2}b_2 + \frac{1}{2} f_2(\lambda b_2)\cdot r/v]\cdot(1-p) \\ &= \frac{1}{2}b_2 + b_2 p + \frac{1}{2} f_2(\lambda b_2)\cdot r/v \equiv \frac{1}{2}b_2 + b_2 p + \frac{1}{2} f_2(\lambda b_2)\cdot s \end{aligned} \tag{3.7}$$

4.1 Performance evaluation

For a policy to be feasible, the bucket size must be greater than or equal to the tour time. Otherwise, couriers would not be able to make it back to the depot before the commencement of the next tour. The following relationship must hold:

$$b_i \geq L_i / v = f_i(\lambda b_i)s \qquad i = 1,2 \tag{4.1}$$

Alternatively, the relationship can be expressed as $b_i = f_i(\lambda b_i)s + \varepsilon$, where ε is the allowance and ε ≥0. By substituting $f_1(\lambda b_1)$ to Equation (4.1), it is obtained that

$$b_1 \geq a_1 \lambda^{c_1} b_1^{c_1} s\,.$$

After re-arranging terms, a feasible domain for b_1 is given by:

$$b_1 \geq (a_1 \lambda^{c_1} s)^{1/(1-c_1)} \equiv \underline{b}_1 \tag{4.2}$$

Similarly, for policy 2, $b_2 \geq (a_2 \lambda^{c_2} s)^{1/(1-c_2)} \equiv \underline{b}_2$ (4.3)

Since both L_i and T_i are increasing functions of the time bucket b, it is optimal to choose the minimum $\underline{b}_i$ for the bucket size decision variable b_i.

Result 3. (*performance comparison*) *Policy 2 will outperform policy 1 in average sojourn time when the job arrival rate* λ *is greater than the threshold*

$$\tilde{\lambda} = (1+p)^{\frac{(1-c_1)(1-c_2)}{c_1-c_2}} a_1^{\frac{c_2-1}{c_1-c_2}} a_2^{\frac{1-c_1}{c_1-c_2}} \cdot s$$

Proof. From Equation (4.1), the feasible sojourn time (Equations 3.6 and 3.7) can be written as:

$$T_1(b_1) = \frac{1}{2}b_1 + \frac{1}{2} f_1(\lambda b_1)\cdot s = f_1(\lambda b_1)\cdot s + \frac{1}{2}\varepsilon \qquad b_1 \geq \underline{b}_1$$

$$T_2(b_2) = \frac{1}{2}b_2 + b_2 p + \frac{1}{2} f_2(\lambda b_2)\cdot s = (1+p) f_2(\lambda b_2)\cdot s + \frac{1}{2}\varepsilon \qquad b_2 \geq \underline{b}_2$$

Both functions are monotonically increasing in the bucket size. For the same batch size q, $f_1(q)$ is at a level higher than $f_2(q)$. However, when different bucket sizes are chosen for the two policies, $T_1(b_1)$ and $T_2(b_2)$ will intersect, owing to the multiplier 1+p in $T_2(b_2)$. (This relationship will be illustrated with an example shortly.) Assume zero allowance ε=0 without loss of generality. For any given λ and p, the minimum sojourn time is obtained at $\underline{b}_i$. By substituting $\underline{b}_1$ and $\underline{b}_2$,

$$T_1\big|_{b=\underline{b}_1} = f_1(\lambda \underline{b}_1)\cdot s = a_1\lambda^{c_1}(a_1\lambda^{c_1}s)^{\frac{c_1}{1-c_1}}s = (a_1\lambda^{c_1}s)^{\frac{1}{1-c_1}}$$

$$T_2\big|_{b=\underline{b}_2} = (1+p)f_2(\lambda \underline{b}_2)\cdot s = (1+p)(a_2\lambda^{c_2}s)^{\frac{1}{1-c_2}}$$

Both functions are monotonically increasing in the arrival rate λ. They intersects at a certain value of λ. For policy 2 to have a lower sojourn time than policy 1, that is, $T_2(\underline{b}_2) \le T_1(\underline{b}_1)$, it is required that

$$\lambda > (1+p)^{\frac{(1-c_1)(1-c_2)}{c_1-c_2}} a_1^{\frac{c_2-1}{c_1-c_2}} a_2^{\frac{1-c_1}{c_1-c_2}} \cdot s \tag{4.4}$$

The right-hand side stipulates a threshold for λ. Denote it as $\tilde{\lambda}$. □

Result 3 states that policy 2 would outperform policy 1 in sojourn time when the arrival rate is greater than $\tilde{\lambda}$. With a large arrival rate, policy 1 will necessitates a large time bucket. The average sojourn time has two components: time bucket size and average tour time. Although the delivery of cross-region jobs is postponed to the next period, reduction in courier workload will have a positive effect on shortening the time bucket. Postponement applies only to *CR* jobs, but shortened time bucket will benefit all jobs. Overall, Result 3 shows that the positive effect of shortened bucket size outweighs the negative effect of postponement when $\lambda \ge \tilde{\lambda}$.

Following policy 2, the workload for couriers will always be smaller since couriers do not make excursion trips across regions. However, policy 2 calls for flexible bucket size. We will next illustrate a procedure for determining the bucket size by analyzing both workload and sojourn time by using Figure 4.

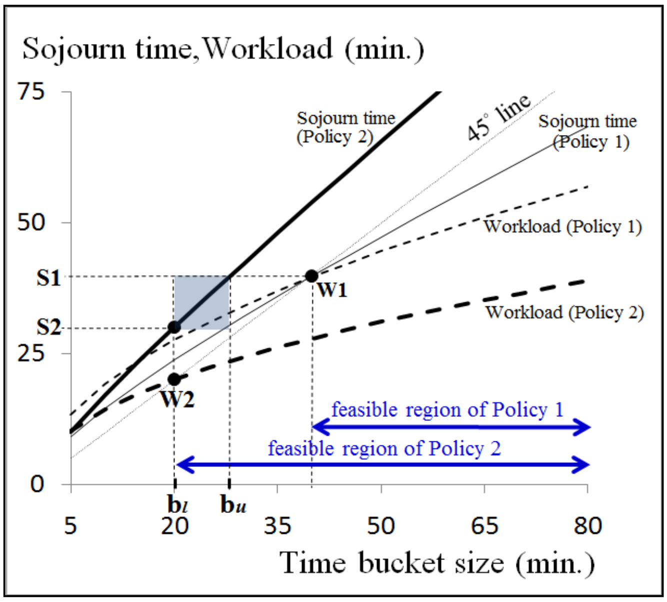

Figure 4: Performance of policy 2.

Figure 4 is obtained for the parameter values of r=1, λ=0.2, p=0.5 and v=0.3. It can be seen that the workload under policy 2 is always smaller than the workload under policy 1 for all bucket size settings (dotted curves). In the figure, the intersection of a workload curve and the 45° line determines a smallest feasible value for a time bucket size. Therefore, the points W1 and W2 demarcate the lower limits of feasible time bucket size for policies 1 and 2, respectively. The corresponding sojourn times are S1 and S2, by reading off the sojourn time functions.

In Figure 4, the point W2 gives a lower limit at b_l for the time bucket. An upper limit b_u is determined from the sojourn time function by using the sojourn time of S1 as the functional value. The rectangular shaded area is constructed from S1, S2, b_l and b_u. Based on the relative location of the sojourn time curves, two cases can be distinguished. In case S2<S1, as shown in Figure 4, the shaded area exists. Then policy 2 is feasible. By choosing any time bucket size in the feasible domain [b_l, b_u], one will obtain a lower workload and a shorter sojourn time simultaneously. Both sojourn time and workload are improved. In contrast, the shaded area does not exist if S2≥S1. In this case, not shown in the figure, policy 2 does not produce a shorter sojourn time.

4.2 Sensitivity analysis

In the above analysis, the performance model contains three parameters: job arrival rate (λ), normalized territory size (s=r/v), and the fraction of *CR* jobs (p). The effects of the three parameters are analyzed in this section. Since it is optimal to choose the smallest feasible bucket size $\underline{b}_i$, take T_1 and T_2 at their lower limits, the ratio $T_2/\ T_1$ can be evaluated as:

$$\frac{T_2\big|_{b=\underline{b}_2}}{T_1\big|_{b=\underline{b}_1}} = (1+p)a_1^{\frac{-1}{1-c_1}} a_2^{\frac{1}{1-c_2}} (\lambda \cdot s)^{\frac{c_2-c_1}{(1-c_1)(1-c_2)}} \tag{4.5}$$

From the property of the tour length functions, we know that $1-c_1>0, 1-c_2>0$, and $c_2-c_1<0$. Therefore, the exponent $(c_2-c_1)(1-c_1)^{-1}(1-c_2)^{-1}$ is negative. The ratio T_2/T_1 is decreasing in λ and r/v. The improvement in percentage of T_2 over T_1 increases with job arrival rate λ and territory size r/v.

*Illustrative Example #*1: In this example, styled data of the hospital case is used to gain insights on the arrival rate threshold and to illustrate the relationship between $T_1(b_1)$ and $T_2(b_2)$. Using the regression formulas of Equations (3.2) and (3.3) and by setting r=100 m and v=50 (m/min), it can be calculated that the threshold $\tilde{\lambda}$= 0.00082, 0.06757 and 0.16333 respectively, for p=0.1, 0.5 and 1.0. The threshold is increasing in p. The sojourn times for the two policies under three values of p are plotted in Figure 5.

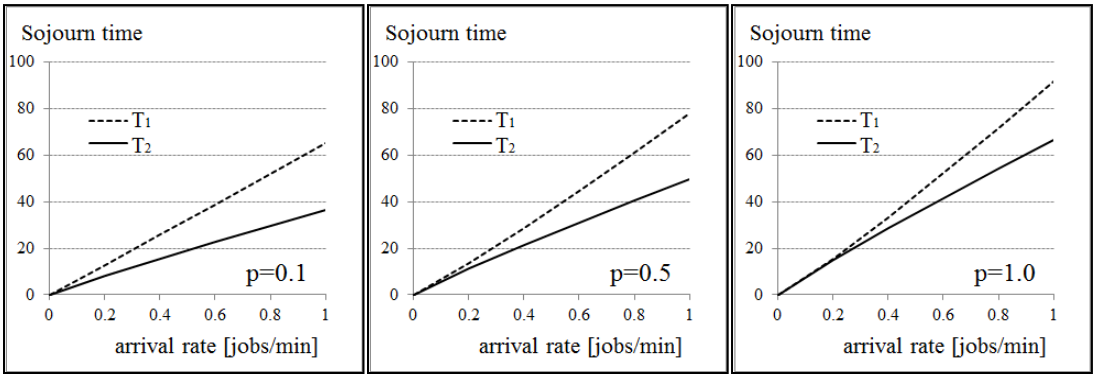

Figure 5: Effect of arrival rate and *CR* job fraction on sojourn time.

Figure 5 has some implications on the applicability of policy 2. The advantage of policy 2, manifested as the gap between T_2 and T_1 or in the ratio T_2/T_1. For any given p, the gap is increasing in the arrival rate. Also, from Equation (4.5), the ratio T_2/T_1 is decreasing in λ and r/v for the same p. The normalized territory size r/v has the same effect as λ in the ratio T_2/T_1. Furthermore, the territory size significantly increases the workload under policy 1. Therefore, the advantage of policy 2 is more significant when the arrival rate and territory size are large.

The effect of p on T_2/T_1 is more complicated to analyze than the effect of λ and r/v, since the coefficients a_1 and c_1 are dependent on p (in Equation 4.5). The advantage of policy 2 is attributable to eliminating the need to make lengthy excursions for delivering *CR* jobs. However, when p is increased, there would be more excursion jobs and the density of excursion jobs per unit tour length would increase. As p increases, policy 1 would gradually gain the benefits of the economies of scale. In Figure 5, the ratio T_2/T_1 equals to 0.561, 0.637 and 0.725 for p=0.1, 0.5, and 1.0 respectively. Thus, policy 2 has more potential to improve the sojourn time when p is small.

5. Dynamic Operation

In the previous section, the feasibility and possibility of policy 2 is proven based on mean-value analysis. In this section, its operation execution is evaluated in a dynamic setting by using simulation. A continuous stream of jobs is generated based on a Poisson process for the 3-circle configuration of Figure 1 by using r=1, p=0.5, v=0.3 and λ= 0.6 (λ =0.2 for each region). These parameter values are the same as what is used in Figure 4. The lower limit in bucket size, b_l, is given by $\underline{b}_2$. From Equation (4.3),

$$b_l = \underline{b}_2 = (a_2 \lambda^{c_2} s)^{1/(1-c_2)} = 20.06$$

The upper limit b_u is obtained from Equations (3.6) and (3.7) by solving the equation

$$T_2\big|_{b=b_u} = S1$$

where S1 is the sojourn time of policy 1 at the bucket size b such that $T_1\big|_b = b$. The calculation shows that S1=39.30, S2=30.09 and b_u=27.61. From Result 3, the arrival rate threshold is $\tilde{\lambda} = 0.040$. These calculation results show that the range [b_l, b_u] is not empty; the opportunity for improvement exists.

Each job has an arrival time. Since policy 2 involves transshipment across time periods, the simulation covers jobs in two time periods, i.e., t=0 and 1. For each courier, there are three sets of jobs to consider.

- SR_t: the set of same-region jobs in time t.
- $CR_{t\text{-}1}$: the set of cross-regions jobs that are not delivered in time t-1. For each courier, this set contains all transshipped jobs from other couriers.
- CR_t: Cross-region jobs which are picked up in time t but delivered in time t+1.

This simulation is focused on period 1. The job input for policy 1 is the set $SR_1 \cup CR_1$. The job input for policy 2 is the three sets CR_0, SR_1 and the pickup of CR_1. To compare the performance of the two policies a mixed Integer Linear Programming (MILP) model is used. The optimal route for policy 1 is solved by using the MILP model. For policy 2, an additional constrained is added: the set of CR_0 jobs are delivered before all other P&D jobs of period 1. Since jobs in CR_0 are from a previous time period (arriving in time period -1) and picked up in time 0, they have spent substantial time in the system. By imposing this constraint, policy 2 is actually not given any advantage and performance comparison would not favor policy 2. This constraint is added because it is more acceptable in practical application.

The MILP model for P&D routing is described next, followed by output analysis. Let P be the set of pick-up nodes and D be the set of delivery nodes. Let S and E be the singleton sets for the starting point and ending node. Let N be the set of all nodes, i.e., $N=P\cup D\cup S\cup E$. The total number of pick-up nodes is n and the total number of delivery nodes is m. The nodes are numbered in the following sequence:

$P=\{1, 2, \ldots, n\}$; $D=\{n+1, n+2, \ldots, n+m\}$; $S=\{n+m+1\}$; $E=\{n+m+2\}$

The binary decision variable $z_{i,j}$ indicates an arc between nodes i and j and $c_{i,j}$ is its cost. The objective function is used to minimize the total tour time (or cost). The variable s_i is used to instantiate the service sequence at each node. Let K be the total number of nodes in this routing problem and $K=n+m+2$.

$$Minimize \quad C = \sum_{i,j\in N, i\neq j} c_{i,j} z_{i,j}$$

s.t. $\sum_{i \in P \cup D \cup E} z_{i,n+m+1} = 0$ (Start from node n+m+1) (5.1)

$\sum_{i \in P \cup D \cup S} z_{n+m+2,i} = 0$ (End with node n+m+2) (5.2)

$\sum_{j \in P \cup D \cup S} z_{j,i} = 1 \qquad i \in P \cup D \cup E$ (Come to node i) (5.3)

$\sum_{j \in P \cup D \cup E} z_{i,j} = 1 \qquad i \in P \cup D \cup S$ (Depart from node i) (5.4)

$s_j \geq s_i + z_{i,j} - (K-1)(1-z_{i,j}) + (K-2)z_{j,i} \quad i \neq j$ (5.5)

$s_i \geq s_{p(i)} \qquad i \in D$ (Precedence constraint) (5.6)

$s_{n+m+1} = 1, \qquad s_{n+m+2} = n+m+2$ (Depot as the first and last node) (5.7)

$z_{n+m+1,n+m+2} = z_{n+m+2,n+m+1} = 0$ (5.8)

$z_{i,j} \in \{0, 1\}, \quad z_{i,i} = 0 \qquad i, j \in P \cup D \cup S \cup E$

In Equation (5.6), the functional notation $p(i)$ denotes the pickup node of a delivery node i. Equation (5.5) enforces sequential relationships between two nodes i and j. Its logic is explained as follows. There are three mutually exclusive cases for the relationship between any two nodes i and j in a route: (1) i precedes j, (2) j precedes i, and (3) i and j are not adjacent. The three cases are also exhaustive. Equation (5.5) governs all three cases. When node i precedes node j, $z_{i,j}$=1 and $z_{j,i}$=0. The constraint reduces to $s_j \geq s_i$+1. When j precedes i, $z_{i,j}$=0 and $z_{j,i}$=1. The constraint reduces to s_j=s_i-1. When i and j are not adjacent, $z_{i,j}$=0, $z_{j,i}$=0 and the value of s_i and s_j should be unrelated. The constraint reduces to $s_j \geq s_i$-(K-1). Since the right-hand side is less than or equal to 1 and the left-hand side is greater than or equal to 1, this reduced form is redundant; it enforces no specific relationships between s_i and s_j. Finally, it should be noted that Equation (5.5) is linear. After the optimal sequence is determined, the arrival time at each node can be easily calculated.

The simulation experiment on dynamic operation has two phases. In the first phase, a fixed time bucket of 40 is used. The performance of both policies is compared. In the second phase, the bucket size is varied for policy 2 to improve both performance measures of utilization and sojourn time. The input and output of phase one is listed in Table 2. The number of jobs in SR_t, CR_t and CR_{t-1} for each courier is shown at the top. The workload and sojourn time by courier is shown in the bottom. In dynamic operation, the 40 minute bucket is feasible for both policies, since all couriers will be able to return to the depot within 40 minutes (with workload of 30.576, 27.708, and 39.294 respectively for the three couriers, for example under policy 1). It can be seen that the average sojourn time is higher for policy 2. Thus, policy 2 is inferior to policy 1.

Table 2: Output of dynamic operation with fixed time bucket.

	Courier	P&D order size			Bucket size
		SR_t	CR_t	CR_{t-1}	
Input	A	3	3	2	40
	B	2	3	5	
	C	3	4	4	
Ouput	Courier	Policy 1		Policy 2	
		Workload	Sojourn time	Workload	Sojourn time
	A	30.576	40.486	30.246	48.163
	B	27.708	34.165	29.619	54.723
	C	39.294	37.727	28.481	59.621

In a second phase of experiment, the time bucket is reduced from 40 in steps of 5 minutes. The same job stream is used. When the bucket is changed, the job sets for input are re-compiled based on their arrival times. The results are shown in Table 3. In operation, we can see that the time buckets of 35, 30, 25 and 20 are all infeasible for policy 1, as some couriers have workload exceeding the bucket size (highlighted in gray background). Those couriers would not be able to make it back to the depot in time before the commencement of the next service trip. The situation for policy 2 is different. The time buckets of 35, 30 and 25 are feasible. The sojourn times decrease with the bucket size. By setting the time bucket at 25, policy 2 will produce a smaller average sojourn time than policy 1 at time bucket of 40. The courier utilization is 0.813 vs. 0.807 for policies 1 and 2, respectively. For policy 1, the average workload per courier is 32.526 (which is calculated from 30.576, 27.708 and 39.294) and hence courier utilization is 0.813 (which is calculated from 32.526 and bucket 40). Courier workload is also lower under policy 2 than policy 1. Thus, this experiment demonstrates that the policy of transshipment across flexible time period can improve both performance criteria.

Table 3: Output with varied time bucket sizes.

Bucket size	Courier	Policy 1		Policy 2	
		Workload	Sojourn time	Workload	Sojourn time
35	A	30.370	40.518	29.653	48.335
	B	27.708	29.165	29.536	45.631
	C	35.798	33.952	22.391	47.949
30	A	30.370	35.518	29.653	43.335
	B	23.713	24.803	28.228	40.195
	C	25.027	29.155	19.040	35.551
25	A	27.675	34.222	22.046	32.613
	B	18.774	19.030	21.429	36.516
	C	25.027	24.155	17.020	22.523
20	A	27.675	29.222	22.046	27.613
	B	8.791	9.964	14.897	25.441
	C	24.577	21.354	15.309	20.187

6. Discussion and Conclusion

In routing problems, tradeoffs between multiple conflicting objectives are fundamental. It is common in the literature to employ additive utility functions or goal programming. In this paper, the approach of designing operational policy is taken. A salient policy (policy 2) based on hub transshipment across flexible time bucket is proposed. The advantage of the policy is analyzed by using stochastic routing analysis and mathematical derivation. Numerical calculation based on styled data demonstrates that policy 2 can be superior to the periodical policy (policy 1) when the service territory is large and the arrival rate is high. When the service territory is large, lengthy excursion trips are detrimental to service quality. Policy 2 overcomes this peril by restricting each courier to work in a dedicated sub-region. By clustering jobs across periods, policy 2 further reaps the benefits of economies of scale. Through efficient workload aggregation and, thus, reduced time bucket, policy 2 leads to a reduction in sojourn time for *all* jobs which outweighs increased sojourn time for *CR* jobs. Therefore, the *CR* job fraction is a critical factor (which is embodied in Result 3). These results on job clustering and workload aggregation are a refinement of general knowledge on the effect of high demand on routing efficiency. The flexible time bucket policy has another advantage if the total demand volume changes over time or over different times of the day. The length of time periods can be dynamically adjusted to changing demand load. This can be a direction of future research.

The square root formulas for tour length that have been reported in the literature for other problems are shown to hold in the problem settings of P&D routing with cross-region routing. This contribution points to a wider applicability of the formulas in general vehicle routing problems.

Acknowledgements

This study is partially supported by funding from the National Science Council of Taiwan under grant 99-2221-E-002-153-MY3.

References

[1] Angelelli, E., Bianchessi, N., Mansini, R. and Speranza, M. G. (2009). *Short term strategies for a dynamic multi-period vehicle routing problem*, Transportation Research Part C: Emerging Technologies, Vol.17, 106-119.

[2] Athanasopoulos, T. and Minis, I. (2011). *Multi-period routing in hybrid courier operations*. In: Minis, I., Zeimpekis, V., Dounias, G. and Ampazis, N. (EDs), Supply Chain Optimization, Design, and Management: Advances and Intelligent Methods, ICI Global, New York, 232-251.

[3] Berbeglia, G., Cordeau, J. F. and Laporte, G. (2010). *Dynamic pickup and delivery problems*, European Journal of Operational Research, Vol.202, 8-15.

[4] Cortés, C. E., Matamala, M. and Contardo, C. (2010). *The pickup and delivery problem with transfers: Formulation and a branch-and-cut solution method*, European Journal of Operational Research, Vol.200, 711-724.

[5] Daganzo, C. F. (1984). *The length of tours in zones of different shapes*, Transportation Research Part B: Methodological, Vol.18, 135-145.

[6] Gent, I. P. and Walsh, T. (1996). *The TSP phase transition*, Artificial Intelligent, Vol.88, 105-109.

[7] Larsen, A., Madsen, O. and Solomon, M. (2002). *Partially dynamic vehicle routing-models and algorithms*, The Journal of the Operational Research Society, Vol.53, 637-646.

[8] Lin, C. K. Y. (2008). *A cooperative strategy for a vehicle routing problem with pickup and delivery time windows*, Computers & Industrial Engineering, Vol.55, 766–782.

[9] Liu, Q. and Xu, J. P. (2008). *A Study on vehicle routing problem in the delivery of fresh agricultural products under random fuzzy environment*, International Journal of Information and Management Sciences, Vol.19, 673-690.

[10] Mes, M., van der Heijden, M. and Schuur, P. (2010). *Look-ahead strategies for dynamic pickup and delivery problems*, OR Spectrum, Vol.32, 395-421.

[11] Mitrovic-Minic, S. and Laporte, G. (2004). *Waiting strategies for the dynamic pickup and delivery problem with time windows*, Transportation Research Part B: Methodological, Vol.38, 635-655.

[12] Mitrovic-Minic, S. and Laporte, G. (2006). *The pickup and delivery problem with time windows and transshipment*, INFOR Information Systems and Operational Research, Vol.44, 217-227.

[13] Nakao, Y. and Nagamochi, H. (2008). *Worst case analysis for pickup and delivery problems with transfer*, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E91-A, 2328-2334.

[14] Parragh, S. N., Doerner, K. F. and Hartl, R. F. (2008). *A survey on pickup and delivery problems Part II: Transportation between pickup and delivery locations*, Journal für Betriebswirtschaft, Vol.58, 81-117.

[15] Pureza, V. and Laporte, G. (2008). *Waiting and buffering strategies for the dynamic pickup and delivery problem with time windows*, INFOR Information Systems and Operational Research, Vol.46, 165-175.

[16] Swihart, M. R. and Papastavrou, J. D. (1999). *A stochastic and dynamic model*

for the single-vehicle pick-up and delivery problem, European Journal of Operational Research, Vol.114, 447-464.

[17] Wen, M., Cordeau, J. F., Laporte, G. and Larsen, J. (2010). *The dynamic multi-period vehicle routing problem*, Computers & Operations Research, Vol.37, 1615-1623.

[18] Winch, J. K., Madu, C. N. and Kuei, C. H. (2012). *Metamodeling and optimizing a reverse logistics system*, International Journal of Information and Management Sciences, Vol.23, 41-58.

Department of Mechanical Engineering, National Taiwan University, Taiwan, ROC
Email: hunter961016@gmail.com
Major areas: Logistics, production and operation management, quality management

Department of Mechanical Engineering, National Taiwan University, Taiwan, ROC
Institute of Industrial Engineering, National Taiwan University, Taiwan, ROC
Email: ychou@ntu.edu.tw
Major areas: Supply chain management, production and operation management, industrial economics.