Netw Spat Econ (2006) 6: 221–234
DOI 10.1007/s11067-006-9281-z

Optimal Fleet Allocation of Freeway Service Patrols

Yafeng Yin

Abstract As one component of traffic incident management systems, freeway service patrols (FSP) facilitate quick removal of incidents through faster response and reduced clearance time. This paper is to investigate how to allocate tow trucks among patrol beats to maximize the effectiveness of the FSP services. A min–max bi-level programming model is proposed to determine an optimal fleet allocation that minimizes the maximal system travel time that incidents may incur. A heuristic iterative solution algorithm is proposed to solve the model. Both the model and the algorithm are demonstrated and validated through a numerical example.

Keywords Freeway service patrols · Fleet allocation · Robust optimization

1 Background

Traffic incident management is a planned and coordinated process to detect, respond to, and remove traffic incidents and restore traffic capacity as safely and quickly as possible. It has emerged as a proven solution to ensure highway efficiency and reliability (Farradyne, 2000). As one component of incident management systems, freeway service patrols (FSP) facilitate quick removal of incidents through faster response and reduced clearance time.

FSP typically operate as follows. The freeways are divided into disjoint beats, each 10–20 miles long with a certain number of tow trucks patrolling on. These trucks travel back and forth along the beat, stopping to clear incidents in a first-reach-first-serve manner. The tow trucks would remove the vehicles stalled in the freeways and provide services such as changing flat tires and offering a needed gallon of gasoline. If they cannot get the vehicles operational in a few minutes they will tow them off the freeway to a designated area (Petty, 1997). FSP systems have been deployed extensively across the U.S., such as in Chicago, Los Angeles and the San Francisco Bay Area. Reviews on these practices can be found in Morris and Lee (1994), Fenno and

Y. Yin (✉)
Department of Civil and Coastal Engineering, University of Florida,
365 Weil Hall, P.O. Box 116580, Gainesville, FL 32611, USA
e-mail: yafeng@ce.ufl.edu

Odgen (1998) and Skabardonis et al. (1998) among others. Note that the way FSP systems operate is different from that of incident-response dispatch systems. FSP tow trucks spontaneously detect, respond to and clear the incidents. In contrast, in the incident-response systems trucks are placed at certain depots, waiting for the dispatch commands. Once an incident is detected or reported, the dispatch center will dispatch a truck to the incident location.

Previous research has examined the benefits of FSP. For example, Skabardonis et al. (1998) evaluated the FSP system on a 7.8 mile section of I-10 freeway (Beat 8) in Los Angeles, and reported that the services reduce incident duration in the order of 15 min and the B/C ratio is greater than 5 where benefits calculated include delay and fuel savings. Levinson et al. (2003) carried out a stated preference analysis to investigate the utility that FSP provide to an individual and found that the B/C ratio for the Los Angeles FSP is in the range of 6.2–6.3. Moore II et al. (2004) examined the prevailing assumption that FSP may reduce the likelihood of secondary accidents and concluded that secondary accidents on Los Angeles freeway are much less frequent than generally reported and avoiding secondary accidents provides only a small incentive to deploy FSP. However, the expected benefits associated with reducing already low secondary rates may be sufficient to justify the program.

It has been well recognized that FSP deployment strategy is the key to the success of the program. Design of the strategy involves determination of patrol beats, fleet size, allocation of the fleet among beats and hours of operations etc. In practice, these decisions are often made based upon engineering experience and judgment. In view of this, investigations have been conducted to develop simulation, statistical and optimization models to help with the decision makings.

Pal and Sinha (2002) and Ozbay and Bartin (2003) have developed simulation models for evaluating various FSP system configurations. Certainly, if a small number of alternatives can be predetermined, such simulation models can be adopted to select the best deployment or expansion strategy of FSP services.

Davies et al. (2004) developed a tool to determine the B/C ratio for providing new FSP service to a freeway section or enhancing the existing service. Given the number of tow trucks on the section, the tool uses statistical models, derived from analysis of over 120 existing beats with 680,000 assists in California, to estimate the delay, fuel and emission savings per assisted incident, and consequently calculate the B/C ratio. The tool is helpful to the decision making on where to implement the next service patrol. Khattak et al. (2004) developed another tool for the same purpose. The tool allows users to obtain statewide rankings for a freeway section based on three index criteria, and estimate the B/C ratio of implementing FSP on that section. The estimates are made primarily based upon statistical data and user inputs. It should be pointed out that both decision-support tools focus on the facility level and localized impacts, lacking a systems perspective.

This paper does not attempt to address all of the issues associated with FSP deployment strategies. The paper is only concerned with how to assign tow trucks to FSP beats to maximize the effectiveness of FSP services, given the setup of the beats and the fleet size. In the current practices, the allocation is made in a heuristic manner. Uniform allocation is sometimes adopted, or the fleet is allocated proportionally to the criteria like traffic volume, vehicle miles traveled and incident rates etc.

The most directly relevant research from the literature is Petty (1997) and Ozbay et al. (2004). Petty (1997) proposed a model for determining where to place tow trucks so as to maximize the expected reduction in congestion, based on traffic theory in combination with marginal benefit analysis. Ozbay et al. (2004) developed a mixed-integer programming model to determine the number of service vehicles assigned to each depot given the locations of depots, and the distribution of incident occurrences. Both studies assume a prior knowledge of incident occurrence distributions, and do not take into account interactions among system performance, incident occurrences, drivers' spontaneous responses to incidents, and service intensities of FSP on various beats.

The remainder of the paper is organized as follows. Section 2 presents the model formulation for FSP fleet allocation problem, and Section 3 proposes a solution algorithm for the model, followed by a numerical example in Section 4 to demonstrate and verify the model formulation and solution algorithm. Conclusions and recommendations for further research are offered in the last section.

2 Model Formulation

2.1 Definition of the Problem

We consider a FSP fleet allocation problem for a general traffic network. Given a limited number of FSP tow trucks and the setup of FSP beats, the decision to make is assigning trucks to FSP beats to maximize the effectiveness of FSP services. Since the length of each beat is fixed, more tow trucks patrolling on the beat imply higher service intensity, quicker removal of incidents and consequently less incident-induced delay. Since the impacts of similar incidents may vary significantly from location to location, where to place the tow trucks is critical to the effectiveness of the services.

There exist a variety of ways to represent the effectiveness of FSP services, such as the expected reduction in congestion. Considering the fact that decision makers are mostly risk averse and are very often more concerned with the high-consequence scenarios, this study attempts to minimize the maximal system travel time that incidents may incur. In other words, we try to obtain a fleet allocation that achieves the least system total travel time under its corresponding worst-case scenario caused by incidents (note that the worst-case scenarios could be different for different fleet allocation plans).

2.2 Time-Independent Modeling Framework

A time-independent modeling framework is used in this paper. There are two reasons for adopting such a static framework. Firstly, as aforementioned, FSP work in a way different from the real-time dispatch systems, and the decision of FSP fleet allocation is "once for all" rather than real time. Certainly different patterns of service intensity or vehicle allocation could be determined for different times of day. Secondly, for each feasible vehicle allocation it is favorable to use a time-dependent analytical or simulation traffic model to evaluate effectiveness of that alternative. However, such a model is always too complicated to be incorporated in the optimization procedure. Moreover, since the model to be developed in this paper is intended for the planning purpose, details of traffic dynamics may not be a major concern at such a macroscopic level.

2.3 Basic Settings

Consider a network $G = (N, A)$, where N is the set of nodes, and A is the set of links. Let W be the set of all origin–destination (O–D) pairs in the network, R_w be the set of routes between O–D pair $w \in W$ and q_w be the demand between O–D pair w. Denote the number of FSP beats as I, and B_i is the set of links that beat i comprises. Let f_r^w be the flow on route $r \in R_w$, $w \in W$, and v_a be the traffic flow on link $a \in A$. We thus have the following flow conservation equations:

$$v_a = \sum_{w \in W} \sum_{r \in R_w} f_r^w \delta_{ar}^w, a \in A \tag{1}$$

$$\sum_{r \in R_w} f_r^w = q_w, w \in W \tag{2}$$

$$f_r^w \geq 0, r \in R_w, w \in W \tag{3}$$

where $\delta_{ar}^w = 1$ if route r between O–D pair w uses link a, and 0 otherwise. Denote the travel time for each link $a \in A$ as $t_a(v_a, c_a)$, which is assumed to be an increasing/decreasing and strictly convex function of link flow v_a on that link/the capacity c_a of that link. Consequently, the route travel time is:

$$t_r^w = \sum_{a \in A} t_a(v_a) \delta_{ar}^w, r \in R_w, w \in W \tag{4}$$

where t_r^w is the travel time on route $r \in R_w$ between O–D pair $w \in W$.

An incident may reduce freeway capacity for a certain period of time. With a static modeling framework, to represent the impacts, we assume the capacity of each link varies within a certain range due to incidents. Mathematically, for each link a we have:

$$c_a^0 - \varepsilon_a c_a^0 \leq c_a \leq c_a^0 + \varepsilon_a c_a^0 \tag{5}$$

where c_a^0 is the nominal link capacity. The nominal capacity could be the design capacity, and it is not uncommon that actual freeway throughput is greater than this nominal capacity for a certain period of time. ε_a is the coefficient of link capacity uncertainty (variability), whose value depends on the characteristics of that link, such as frequency and severity of incidents, and geometry. The value can be calibrated using historical incident data including locations, types and durations. A two-step procedure could be used for the parameter calibration. The first step is to determine the magnitude of capacity reduction according to the type and location of an incident, e.g., using values suggested in Highway Capacity Manual (TRB, 2000) while the second step is to compute the percentage of time when the capacity is reduced in the peak hour, considering the duration of the incident. Consequently, we obtain the value of ε_a^i for incident i. Going through the two-step procedure for all of the recorded incidents, the value of ε_a can thus be quantified as the maximum of ε_a^i.

If the capacity of each link varies independently, the uncertainty set of link capacity pattern for the whole network will be a box. Recall that we are concerned with the worst case incurred by incidents. With a box uncertainty set, if Braess' paradox (Braess, 1968) is not present, it is straightforward to identify the worst-case scenario where each link has its minimal capacity. However, it is rare, if not impossible, that such a case would ever occur in reality. In other words, the box uncertainty set is too conservative. To be more realistic, we define an ellipsoidal uncertainty set to confine the link capacity pattern for a general network, written as:

$$C = \left\{ c \in R^{|A|} \middle| \sum_{a=1}^{|A|} \left(\varepsilon_a c_a^0\right)^{-2} \left(c_a - c_a^0\right)^2 \leq 1 \right\} \tag{6}$$

where $|A|$ is the dimension of set A (total number of links). The set can be also written as: $C = \left\{ c \in R^{|A|} \middle| c = c^0 + M \cdot u, \|u\|_2 \leq 1 \right\}$, where M is a diagonal matrix whose element is $\varepsilon_a c_a^0$. Note that such an ellipsoidal uncertainty has been widely used in a recent stream of research on robust optimization (e.g., Ben-Tal and Nemirovski, 2002 and El Ghaoui, 2003). The justifications for using such an ellipsoidal uncertainty here are given as follows:

- The ellipsoidal set results in a computationally-tractable model formulation.
- The set is not too conservative because no element in the set implies that capacities of all links achieve their respective minimum simultaneously.
- The set may be given parametrically by observation data of moderate size. As aforementioned using the incident data, the value of ε_a can be calibrated. Consequently the set can be created mathematically. Note that with the min–max notion employed in this paper, we do not intend to incorporate all the possible realizations of link capacities into the uncertainty set. Indeed, it has been shown that even though the uncertainty set does not contain a single realization of the random vector, the min–max concept still results in a meaningful robust solution (Ben-Tal and Nemirovski, 1999).

2.4 Representation of FSP Impacts

Since FSP trucks continuously patrol on beats looking for incidents to assist, they would be able to respond to incidents more quickly and thus reduce incident durations. Incidents may reduce freeway capacities, and consequently locations of incidents could become bottlenecks. Essentially, FSP are able to reduce durations of activation of the bottlenecks. With the time-independent modeling framework, it is impossible to exactly replicate such impacts of FSP. Rather, we assume that FSP may shorten the range of capacity variation of the link that trucks are patrolling on. In other words, the impacts of the services are represented as reducing the variability of link capacity.

With an ellipsoidal uncertainty set of link capacity of the network, what FSP services really change is the geometry of the set, as illustrated in Fig. 1 for an illustrative example with a two-link network, where the outer ellipse represents the uncertainty set before FSP while the inner ellipse is the set after implementing FSP at link 1. Note that, different FSP fleet allocations may change the uncertainty set differently, leading to different worst-case scenarios realized from the corresponding sets. More precisely, the uncertainty set C should be written as $C(z)$, where z is the vector of fleet allocation.

It is easy to know that the intensity or frequency of FSP tow trucks on each beat is given as:

$$h_i = \frac{z_i}{t_i} \tag{7}$$

where h_i is the service intensity at beat i, z_i is the number of tow trucks assigned to beat i, and t_i is the round trip time of beat i. This round trip time is endogenously or exogenously (depending on whether tow trucks allowed to use shoulders or not) calculated by segment travel times, layover times at ends of trips and incident clearance time averaged across the trips made within the time period of interest.

We further represent the relationship between service intensity of FSP and link capacity variability. For $\forall a \in B_i$, the set of links that beat i comprises, we assume the following relationship:

$$\varepsilon_a = s(h_i) = s'(z_i) \tag{8}$$

where s or s' is a decreasing function of service intensity or number of assigned tow trucks. The relationship is intuitively correct, and could be calibrated from empirical data used in Davies et al. (2004) and Dowling et al. (2004). The data were collected from 118 existing FSP beats in California, including information on beat characteristics, average annual daily traffic, number of FSP tow trucks serving

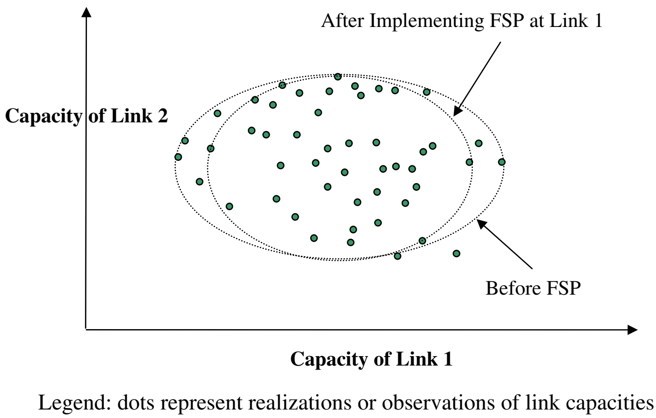

Fig. 1 An illustrative example of FSP impacts

the beat and number of the FSP assists. From the data, a relationship between reduction of incident duration and the FSP service intensity can be established. Consequently, we can obtain a relationship between the percentage of time of capacity reduction in the peak hour and the service intensity. Multiplying the value of ε_a without FSP will lead to the relationship in Eq. (8).

2.5 Estimation of Incident Impacts

As aforementioned, we attempt to determine a fleet allocation that minimizes the maximal system travel time realized from the uncertainty set of link capacities. An element in the set represents a possible realization of uncertain link capacities, incurred by incidents. Intuitively, we may need to evaluate the resulting system performance for each element in the set, and then find the maximum. Note that for any given element or realization, the corresponding system performance is deterministic. However, in order to accurately estimate the performance, we need to describe how drivers react to incidents in a dynamic setting.

Because the modeling framework we propose is intended for planning purposes, we adopt static user equilibrium model (Beckmann et al., 1956) to describe route choice behaviors. For any (deterministic) realization of the uncertain link capacity pattern from the ellipsoidal set, we solve a corresponding network equilibrium problem with the realized link capacities to *approximately* estimate the resulting system travel time incurred by the incidents. We further note that under the static modeling framework, it is feasible to describe drivers' spontaneous responses to incidents in a degradable network, which would lead to a more accurate estimate of system performance. Previous studies have been conducted along this direction, such as Mirchandani and Soroush (1987), Uchida and Iida (1993), Chen et al. (2002), Yin and Ieda (2001), and Lo and Tung (2003) among others. Our future research will look into incorporating such a model.

The formulation of the user equilibrium model is presented as Eqs. (11)–(13), and the definitions of the variables have been given in Section 2.3.

2.6 Formulation

We observe that decision makers tend to be risk-averse and are more concerned with the worst-case scenarios. Therefore, it would be more desirable to have a tow-truck fleet allocation that performs better in the worst-case scenario even though the average performance is relatively poorer. In view of this, we intend to determine a fleet allocation that minimizes the maximal system travel time that incidents may incur. The key point here is that we assume the uncertain capacity caused by incidents is bounded by the ellipsoidal uncertainty set, whose geometry can be further affected by the fleet allocation. By selecting such an ellipsoidal set, we can avoid being overly conservative while the resulting fleet allocation will perform much better against the worst-case scenario without losing much "nominal" optimality.

With the above considerations, the optimization problem can be written as:

$$\min_{z} \max_{c \in C(z)} \sum_{a} v_a t_a(v_a, c_a) \tag{9}$$

subject to:

$$\sum_{i}^{I} z_i \leq Z \tag{10}$$

where v_a is obtained by solving the following lower-level problem:

$$\min \sum_a \int_0^{v_a} t_a(\varpi, c_a) d\varpi \tag{11}$$

subject to:

$$\sum_r f_r^w = d_w, \forall w \tag{12}$$

$$f_r^w \geq 0, \forall r, w \tag{13}$$

where Z is the number of total available trucks. In summary, the upper-level problem also has the following two definition constraints:

$$C(z) = \left\{ c \in R^{|A|} \middle| \sum_{a=1}^{|A|} \left(\varepsilon_a c_a^0\right)^{-2} \left(c_a - c_a^0\right)^2 \leq 1 \right\} \tag{14}$$

$$\varepsilon_a = s'(z_i) \tag{15}$$

and the lower-level problem should satisfy the following definition constraint:

$$v_a = \sum_{w \in W} \sum_{r \in R_w} f_r^w \delta_{a,r}^w, a \in A \tag{16}$$

This is a min–max bi-level programming model. The upper-level problem represents planners' behavior, determining tow truck allocation to minimize the maximal total travel time incurred by incidents. The lower-level problem represents drivers' route choice behaviors, affected by the allocation decision from the upper-level and capacity reductions caused by incidents.

Note that variable z_i, the number of allocated trucks on beat i should be integer, and thus the above model should be an integer programming model. However, due to the computational complexity, this paper treats it as a number. This simplification does not necessarily impair the applicability of the model. In actual applications, one could use the model to obtain optimal service intensity, and then marginally adjust layover times to provide the intensity with an integer number of vehicles, and eventually determine the fleet allocation.

3 Solution Algorithm

The model (9)–(13) is non-convex, and thus only local optima can be found. There is no available solution algorithm for the model. Therefore, we propose a heuristic iterative algorithm to solve the model. Although the algorithm cannot guarantee theoretically convergence to a local optimum solution, it has shown good convergence and led to good results in our numerical experiments.

The iterative algorithm views the model (9)–(13) as a master problem with a slave problem. The master problem is written as:

$$\min_{z} J(z) \tag{17}$$

subject to:

$$\sum_{i}^{I} z_i \leq Z \tag{18}$$

where $J(z)$ is a non-convex function, defined by the optimal objective function of a slave model, which is to identify the worst-case scenario corresponding to the fleet allocation plan z.

For solving the master problem, many efficient algorithms proposed in the literature of operations research could be potentially applied. Since it is difficult to derive analytically the gradient $\nabla_z J(z)$ and only values of $J(z)$ can be made available, we apply iterative descent methods with finite differencing derivatives, such as the sequential quadratic programming algorithm (SQP) by Han (1976).

The slave bi-level programming model defining $J(z)$ is given as below:

$$J(z) = \max_{c} \sum_{a} v_a t_a(v_a, c_a) \tag{19}$$

subject to:

$$c \in C(z) \tag{20}$$

where v_a is obtained by solving the following lower-level problem:

$$\min \sum_{a} \int_{0}^{v_a} t_a(\varpi, c_a) d\varpi \tag{21}$$

subject to:

$$\sum_{r} f_r^w = d_w, \forall w \tag{22}$$

$$f_r^w \geq 0, \forall r, w \tag{23}$$

A number of algorithms can be applied to solve the slave problem, such as those proposed by Yang et al. (1994) and Chiou (2005). In this paper, we apply the sensitivity-analysis-based iterative method by Yang et al. (1994). The basic idea of the algorithm is to formulate local linear approximation of the upper-level objective function using the derivative information from sensitivity analysis for equilibrium flows (e.g., Tobin and Friesz, 1988), and solve the resultant linear programming problems for a descent search direction. Therefore, the algorithm is in fact a sequence of linear approximation to the original problem.

Note that with the ellipsoidal capacity uncertainty set, the local linear approximation to the original bi-level model turns out to be a quadratically constrained program as follows:

$$\max_{\|u\|\leq 1} \nabla_c T^T \cdot M \cdot u \tag{24}$$

where $\nabla_c T$ is the gradient of the total travel time with respect to link capacity, calculated by conducting the sensitivity analysis for user equilibrium flows. This program can be analytically solved; the optimal objective value is $\sqrt{(M \cdot \nabla_c T)^T \cdot (M \cdot \nabla_c T)}$ and the optimal solution is $M \cdot \nabla_c T \Big/ \sqrt{(M \cdot \nabla_c T)^T \cdot (M \cdot \nabla_c T)}$.

In summary, the outer iteration of the iterative procedure applies a SQP algorithm with finite differencing derivatives to solve the master problem. In each iteration, embedded bi-level slave problems are solved by the sensitivity-analysis-based iterative method for specific fleet allocations to identify the corresponding worst-case scenarios.

4 Numerical Example

A numerical example is now presented to illustrate the proposed model. The example road network shown in Fig. 2 has 13 nodes, 19 links and 4 O–D pairs, adopted from Nguyen and Dupuis (1984). The Bureau of Public Road link travel time function was used

$$t_a(v_a) = t_a^0 \left(1 + 0.15 \cdot \left(\frac{v_a}{c_a}\right)^4\right) \tag{25}$$

And the network characteristics and O–D demand are given in Tables 1 and 2, respectively.

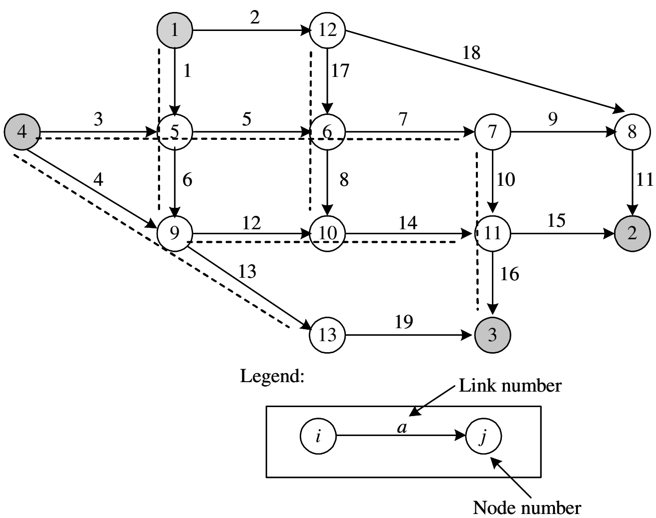

Fig. 2 An example network

Table 1 Network characteristics of the example network

Link a	t_a^0	c_a^0	ε_a^0
1	7.0	800	0.5
2	9.0	400	0.1
3	9.0	200	0.2
4	12.0	800	0.3
5	3.0	350	0.2
6	9.0	400	0.5
7	5.0	800	0.2
8	13.0	250	0.2
9	5.0	250	0.1
10	9.0	300	0.4
11	9.0	550	0.1
12	10.0	550	0.5
13	9.0	600	0.3
14	6.0	700	0.5
15	9.0	500	0.1
16	8.0	300	0.4
17	7.0	200	0.2
18	14.0	400	0.1
19	11.0	600	0.1

In the example, the FSP impact function was assumed to be $s'_a(z) = \varepsilon_a^0 \cdot e^{-0.5^* z_i}$, where ε_a^0 is given in Table 1. The setup of beats is also illustrated in Fig. 2, which is: Beat 1 = {links 1, 6}; Beat 2 = {links 3, 5, 7}; Beat 3 = {links 17, 8}; Beat 4 = {links 12, 14}; Beat 5 = {links 4, 13} and Beat 6 = {links 10, 16}.

A SQP subroutine with finite-differencing derivatives in Matlab was used to solve the master problem (17)–(18) as well as one self-programmed subroutine to solve the bi-level programming model (19)–(23).

We first examined the convergence of the iterative algorithm. Figure 3 plots the value of the objective function against outer iteration number of the SQP procedure, where the size of FSP fleet was 10. It can be observed that the outer iteration of the algorithm had a fast convergence; convergence was achieved in about 15 iterations. However, total computation was quite demanding due to the use of finite-differencing derivatives, which suggests that the iterative algorithm may not be applicable for a large-scale network. The resultant fleet allocation is 3.1, 1.5, 0, 0.7, 1.9 and 2.8 for Beats 1 to 6, respectively.

To validate the effectiveness of the proposed model, we compared system performances that optimal and uniform allocations may result in. The total fleet was set as 6 and uniform allocation means one assigned truck on each beat. We computed differences of total travel times that uniform and optimal allocations could achieve under their corresponding worst-case scenarios. In order to examine the impacts of capacity uncertainty and network congestion, we varied the level of capacity uncertainty by multiplying ε_a^0 listed in Table 1 by an amplifier, changing

Table 2 O–D travel demand for the example network

O/D	2	3
1	400	800
4	600	200

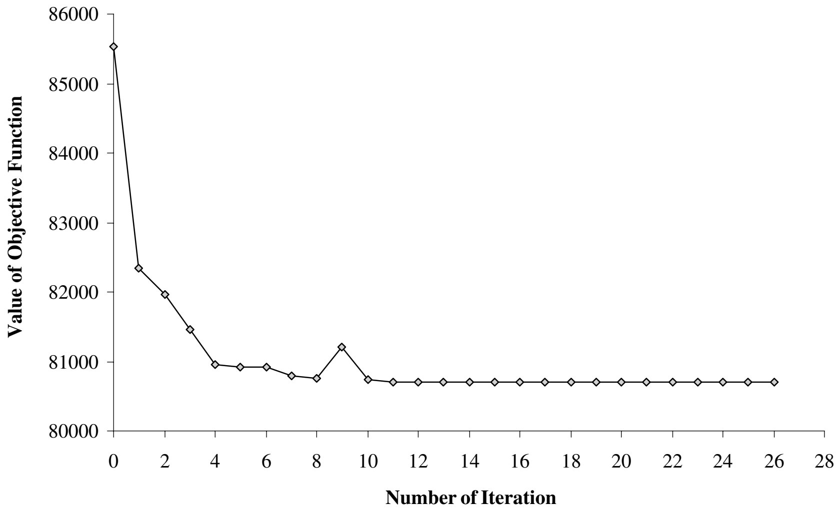

Fig. 3 Convergence of the iterative algorithm for the min–max model

from 0 to 2.0 (value of 0 represents the case of deterministic link capacities), and considered three demand levels: 70, 100 and 110% of the demand given in Table 2.

Figure 4 depicts relative performance differences of optimal and uniform allocations (the worst-case system travel time of uniform allocation minus that of optimal allocation divided by the latter) against varied values of the amplifier for

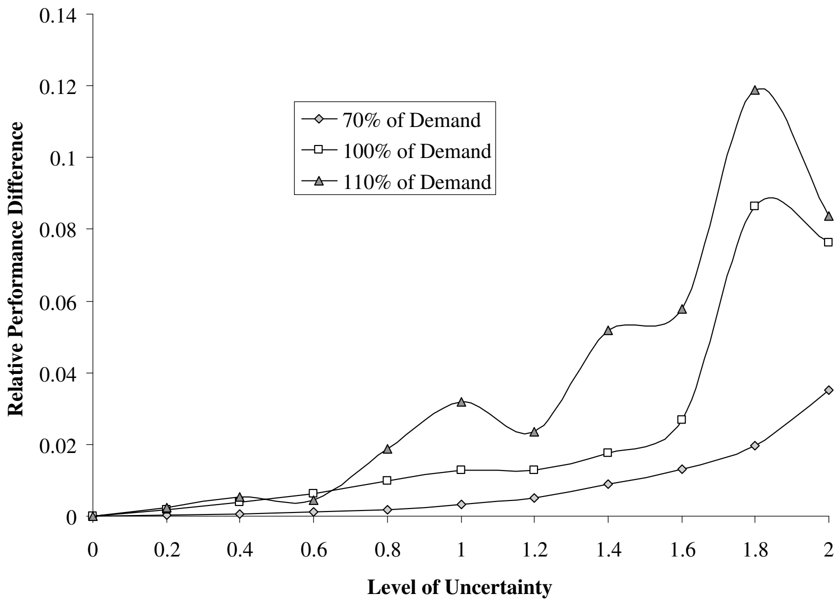

Fig. 4 Relative system performance differences between uniform and optimal allocations (total travel time of uniform minus that of optimal)

three demand levels. The results are within expectation and intuitively correct. We have the following observations from the figure:

- The performance difference is always positive, suggesting that the optimal fleet allocation always performs better against the worst-case scenarios than the uniform allocation does.
- In the case of deterministic link capacities where FSP impose no impacts at all, there is no performance difference between uniform and optimal allocations, as expected.
- The relative performance difference varies from 0 to 12%. Generally speaking, the difference becomes more significant with higher levels of capacity uncertainty or network congestion, suggesting that the model may contribute more or make more difference in the situations of high frequencies of incidents or high levels of network congestion. However, the relative difference that the model can result in may become less significant when the network capacity is highly fluctuating, e.g., the case with the uncertainty amplifier equal to 2.0 in Fig. 4.

5 Concluding Remarks

We have presented a min–max bi-level programming model to determine the optimal fleet allocation strategy for a FSP system. A heuristic solution algorithm has been proposed to solve the model. Both the model formulation and the solution algorithm have been validated and demonstrated through a numerical example.

Further research may follow the following directions: 1) propose more efficient algorithms for the model and apply them to large-scale networks; 2) validate the assumptions of the model formulation and calibrate parameters by using actual data and then apply it to a real-world network; 3) incorporate models that describe drivers' route choice in degradable networks and their impacts on the overall system performance, and 4) extend the model to simultaneously determine the setup of beats and fleet allocation.

Acknowledgments This research was funded by the California Department of Transportation, through a contract, Task Order 5329, to California PATH at University of California at Berkeley, where the author previously worked as an assistant research engineer. The contents of this paper reflect the views of the author, who is responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. The author thanks his previous colleagues at California PATH, Prof. Samer Madanat and Dr. Xiao-Yun Lu for their encouragement and support during the course of the project. The author is also grateful to the anonymous reviewers for their constructive comments.

References

Beckmann MJ, McGuire CB, Winsten CB (1956) Studies in the economics of transportation. Yale University Press, New Haven, Connecticut

Ben-Tal A, Nemirovski A (1999) Robust solutions to uncertain linear programs. Oper Res Lett 25:1–13

Ben-Tal A, Nemirovski A (2002) Robust optimization-methodology and applications. Mathematical Programming, Series B 92:453–380

Braess D (1968) Über ein Paradoxon aus der Verkehrsplanung. Unternehmensforschung 12:258–268
Chen A, Ji Z, Recker W (2002) Travel time reliability with risk sensitive travelers. Transp Res Rec 1783:27–33
Chiou S-W (2005) Bilevel programming for the continuous transport network design problem. Transp Res 39B:361–383
Davies LH, Mauch M, Skabardonis A (2004) Freeway service patrol predictor model: methodology and implementation. In: 83rd Annual Meeting Compendium of Papers CD-ROM. Transportation Research Board, Washington, District of Columbia
Dowling R, Skabardonis A, Carroll M, Wang Z (2004) Methodology for measuring recurrent and non-recurrent traffic congestion. Transp Res Rec 1867:60–68
El Ghaoui L (2003, March 11) Tutorial on robust optimization. Presentation at the Institute of Mathematics and Its Applications. Available at http://robotics.eecs.berkeley.edu/~elghaoui/
Fenno DW, Odgen MA (1998) Freeway service patrols: a state of practice. Transp Res Rec 1643:28–38
Han S-P (1976) Superlinearly convergent variable metric algorithms for general nonlinear programming problems. Math Program 11:263–282
Khattak A, Rouphail N, Monast K, Havel J (2004) Method for priority-ranking and expanding freeway service patrols. Transp Res Rec 1867:1–10
Levinson D, Gillen D, Parthasarathi PK (2003) The insurance value of freeway service patrols: a stated preference analysis. In: 82nd Annual Meeting Compendium of Papers CD-ROM. Transportation Research Board, Washington, District of Columbia
Lo HK, Tung Y-K (2003) Network with degradable links: capacity analysis and design. Transp Res 37B:345–363
Mirchandani P, Soroush H (1987) Generalized traffic equilibrium with probabilistic travel times and perceptions. Transp Sci 21:133–152
Morris M, Lee W (1994) Survey of efforts to evaluate freeway service patrols. Transp Res Rec 1446:77–85
Moore II JE, Giuliano G, Cho S (2004) Secondary accident rates on Los Angeles freeways. J Transp Eng ASCE 130(3):280–285
Nguyen S, Dupuis C (1984) An efficient method for computing traffic equilibria in networks with asymmetric transportation costs. Transp Sci 18:185–202
Ozbay K, Bartin B (2003) Incident management simulation. Simulation 79(2):69–82
Ozbay K, Xiao W, Iyigun C (2004) Probabilistic programming models for response vehicle dispatching and resource allocation in traffic incident management. In: 83rd Annual Meeting Compendium of Papers CD-ROM. Transportation Research Board, Washington, District of Columbia.
Pal R, Sinha K (2002) Simulation model for evaluating and improving effectiveness of freeway service patrol programs. J Transp Eng ASCE 128(4):355–365
Farradyne PB (2000) Traffic incident management handbook. Report USDOT-13286, US DOT
Petty KF (1997) Incidents on the freeway: detection and management. Doctoral dissertation, Department of Electrical Engineering and Computer Science, University of California, Berkeley
Skabardonis A, Petty KF, Varaiya P, Bertini R (1998) Evaluation of freeway service patrol at Los Angeles freeway site. PATH research report, Report UCB-ITS-PRR-98-13, University of California, Berkeley
Tobin RL, Friesz TL (1988) Sensitivity analysis for equilibrium network flow. Transp Sci 22(4)242–250
Transportation Research Board (2000) Highway capacity manual. Washington, District of Columbia
Uchida T, Iida Y (1993) Risk assignment: a new traffic assignment Mmodel considering risk of travel time variation. Proceedings of the 12th international symposium on transportation and traffic theory. Elservier, Amsterdam, 89–105
Yang H, Yagar S, Iida Y, Asakura Y (1994) An algorithm for the inflow control Pproblem on urban freeway networks with user-optimal flows. Transp Res 28B:123–139
Yin Y, Ieda H (2001) Assessing performance reliability of road networks under non-recurrent congestion. Transp Res Rec 1771:148–155