Hongzhi LIN, Ph.D.[1]
E-mail: linhz@seu.edu.cn
School of Economics and Management
Southeast University
Nanjing 211189, China

Intelligent Transport Systems
Original Scientific Paper
Submitted: 5 Nov. 2020
Accepted: 26 Mar. 2021

OPTIMAL ON-RAMP METERING OF URBAN FREEWAY NETWORK FOR THE CORONAVIRUS DISEASE CONTROL

ABSTRACT

The outbreak of COVID-19 disrupted our everyday life. Many local authorities enforced a cordon sanitaire for the protection of sensitive areas. Travellers can only pass the cordon after tested. This paper aims to propose a method to design an on-ramp control scheme to maximise urban freeway network throughput with a predetermined queuing delay constraint at all off-ramps around cordon sanitaire. A bi-level programming model is formulated where the lower-level is a transportation system equilibrium to predict traffic flow, and the upper-level is on-ramp metering optimisation that is nonlinear programming. A stochastic queuing model is used to represent the waiting phenomenon at each off-ramp where testing is conducted, and a heuristic algorithm is designed to solve the proposed bi-level model where a method of successive averages (MSA) is adopted for the lower-level model; A genetic algorithm (GA) with elite strategy is adopted for the upper-level model. An experimental study is conducted to demonstrate the effectiveness of the proposed method and algorithm. The results show that the methods can find a good heuristic optimal solution. These methods are useful for freeway operators to determine the optimal on-ramp control for disease control and prevention.

KEYWORDS

cordon sanitaire; on-ramp control; bi-level programming model; heuristic algorithm; queuing theory.

1. INTRODUCTION

The COVID-19 pandemic is an ongoing global pandemic of a novel coronavirus disease (WHO). The outbreak was first reported in December 2019 in Wuhan, China. Epidemiologists determined that the virus possibly came from an animal sold at a seafood market. However, the source is still not identified. As of 20 October 2020, more than 40.4 million cases of COVID-19 have been reported across 188 countries and territories, resulting in more than 1.1 million deaths. The virus is primarily spread between people during close contact, most often via small droplets produced by coughing, sneezing, and talking. Common symptoms include fever, cough, fatigue, shortness of breath, and loss of sense of smell. Recommended preventive measures include hand washing, covering one's mouth when coughing, maintaining social distance from other people, wearing a face mask in public settings, disinfecting surfaces, increasing ventilation and air filtration indoors, and monitoring and self-isolation for people who suspect they are infected.

In order to contain the rapid spread of COVID-19, authorities worldwide have responded by implementing travel restrictions, lockdowns, workplace hazard controls, and facility closures. Many places have also worked to increase the testing capacity and trace contacts of infected persons. Mostly, a cordon sanitaire was set up on 23 January 2020 to control travel in and out of Wuhan, and then it was extended to many other cities. A cordon sanitaire is the restriction of movement of people travelling to or from a defined geographic area, such as a community, city, or region. Only qualified travellers after testing are allowed to pass the cordon. Such an approach demonstrated to be an effective way to prevent the infectious virus from spreading into a city [1–3]. All travellers passing the cordon are have their body temperature checked to ensure they are not infected. However, it is reported that the queue length is too long and the waiting cost is too high at the cordon sanitaire. Note that the traffic flows at all freeways including off-ramps are determined by the upstream

controlled on-ramp meters. This paper aims to propose a method and an algorithm to design an on-ramp control scheme at an urban freeway network while ensuring a tolerable waiting time at off-ramps around the cordon sanitaire.

As for optimisation, there are many ways to measure the performance of the queuing system around a cordon sanitaire. The most important one is the queuing delay cost, i.e., vehicle waiting time. It is also an important indicator in transport network analysis. Generally speaking, there are two ways to determine queuing delay costs in the transport network. One is proposed by Yang and Yagar [4], where queuing times are determined implicitly. They formulated bi-level programming for on-ramp traffic control problem where the upper-level problem is to determine the ramp metering rates that minimise system total travel time, and the lower-level problem represents a traffic equilibrium model involving ramp queuing. Note that the total travel time is defined as the sum of travel times spent in the corridor and queuing times at controlled ramps by all vehicles. It can be proved that the link queuing time corresponds exactly to the Lagrange multiplier associated with the link capacity constraint. Yang and Yagar [5] further extended the bi-level model to optimise signal timings in saturated road networks. Both queuing and congestion on saturated links are taken into account in predicting equilibrium flows and setting signal split parameters for a fixed pattern of origin-to-destination trip demand. Besides traffic control, Yang and Lam [6] extended the bi-level model for the determination of road toll pattern at bottlenecks, such as tunnels and bridges. The lower-level problem represents a queuing network equilibrium model that describes users' route choice behaviour under conditions of both queuing and congestion. The upper-level problem is to determine road tolls to optimise a given system's performance while considering the route choice behaviour of users. The previous research studies are all under fixed origin-to-destination trip demand so Yang and Bell [7] further extended to an elastic-demand network equilibrium model with queues. The queuing delay time is not explicitly represented in this way. It is determined implicitly as the Lagrange multiplier associated with the link capacity constraint in the lower-level problem. Queues only form when capacity is reached. Otherwise, link travel time will solely depend on flow.

The other approach to determine queuing delay is Vickrey's bottleneck model [8] where the queuing delay time is derived explicitly using the deterministic queuing theory. It aims to address the departure time choices of commuters on a bottleneck-constrained freeway during the morning rush hours. This model is able to model the formulation and dissipation of queuing behind the bottleneck in a tractable and straightforward way, thus making it a benchmark representation of the dynamics of traffic congestion in peak periods. The classical bottleneck model states that the travel time on each link is a free-flow travel time plus a queuing delay before the downstream bottleneck. The bottleneck model research studies discovered a significant progress in the past 50 years. A lot of insights into understanding the features of traffic queuing in peak periods have been obtained via the bottleneck model. Li et al. [9] recently gave a very good bibliometric review and future research directions on the bottleneck models, where interested readers can find much more details.

Although queuing delay at a cordon sanitaire is analogous to that at ramp meters, signalised intersections, and bottlenecks, it has different characteristics that provide potential research opportunities for further studies. First of all, the testing time for each vehicle is stochastic as it depends on the uncertain number of occupants. Therefore, the traditional deterministic queuing model may lead to a large deviation from the actual values and, thus, restricts real applications of the model. In order to model the uncertain testing time at each checkpoint, a stochastic queuing model is specified in this research. Secondly, the aspiration level of queuing delay is highlighted at cordon sanitaire. Although the queuing phenomenon is well explored in the transport networks, the manipulation for the desired level of queuing delay at cordon sanitaire is not investigated to our knowledge. Lastly, the existing studies have mainly focused on the topics of travel behaviour analysis and demand-side strategies (particularly on congestion pricing). However, only limited attention was given to the topics of supply-side strategies [9]. This research aims to propose a novel bi-level programming model and an efficient heuristic algorithm to design on-ramp metering rates of an urban freeway network for epidemic disease control and prevention.

The structure of this paper is organised as follows. Section 2 elaborates on the proposed bi-level programming model. Section 3 describes explicitly a heuristic algorithm designed to solve the proposed model. Section 4 demonstrates the effectiveness of the model and algorithm by means of an experimental study. Section 5 concludes this paper.

2. METHODOLOGY

In order to avoid unbearable waiting time at the testing station, this paper aims to propose a method and an algorithm to design an on-ramp control scheme at an urban freeway network. The traffic flows are indirectly controlled at on-ramps so that they cannot exceed the testing capacity at off-ramps. This is a Stackelberg game with a leader-follower decision structure. The operator in the upper-level aims to maximise network throughput by metering on-ramps. The operator can predict, but cannot control travel behaviours of freeway users including destination choice and route choice, while all users make their own decisions in a user optimal manner. The users' decisions at the lower-level are made after the upper-level decisions. However, the operator must anticipate the behavioural responses of the users to adjust his decisions. A bi-level optimisation model is usually proposed to capture the leader-follower nature of the relationship between the operator and users. The conceptual framework is shown in *Figure 1*. The upper-level is a network throughput maximisation with a queuing delay constraint. The lower-level is a transportation system equilibrium with queuing where travel behaviours of freeway users can be predicted. Note that the equilibrium is a feedback procedure between a trip distribution and traffic assignment with fixed travel demand. The detailed models are elaborated in the following sections.

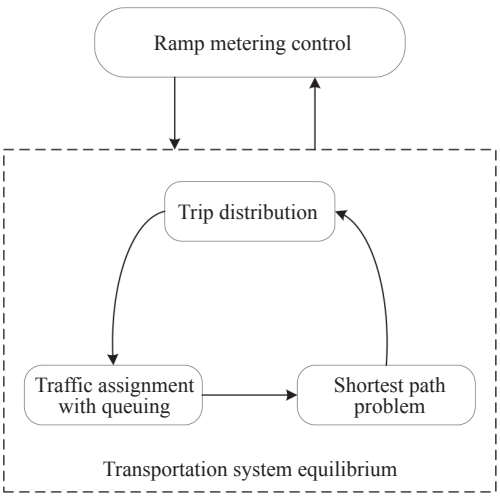

Figure 1 – The conceptual framework of bi-level programming model

2.1 Network equilibrium with queuing

The lower-level model is a transportation system equilibrium that combines trip distribution and traffic assignment models with settled metering rates and given road network. It has long been criticised that travel times are inconsistent in the conventional four-step sequential model because travel times are determined endogenously in fact. Generally speaking, two ways can be used to solve the inconsistent problem to achieve transport system equilibrium according to the literature. One way is to combine several steps to an equivalent mathematical programming, which can achieve a well-converged and consistent result [10, 11]. The other is to feedback the sequential models iteratively until travel times meet the consistency criteria [12, 13]. Although the former is commonly adopted in literature, the latter is more flexible at each step [14, 15]. Therefore, a combined model with feedback is adopted here.

Note that traffic assignment is not a traditional one here as queuing delay at cordon sanitaire is accommodated. The determination of queuing delay time is a critical problem. Generally, queuing theory is an excellent tool to analyse the cost of vehicles waiting. In most traffic situations, interarrival and service times are described randomly by the exponential distribution. This stage adopts a stochastic queuing model that combines both arrivals and departures based on the Poisson assumptions. That is, the interarrival and the service times follow the exponential distribution. In addition, the first-in-first-out (FIFO) discipline is respected on each off-link. The derivation of the specialised queuing model is based on the steady-state behaviour of the queuing situation, achieved after the system has been in operation for a sufficiently long time.

According to the conventional traffic flow theory [16], the waiting line at each toll booth can be formulated as a fundamental $M/M/c$ queuing model, where M means Markovian (or Poisson) arrivals or departures distribution or equivalently exponential interarrival or service time distribution, and c means the number of identical parallel servers with

same service rate per unit time. There could be one or more parallel checkpoints (i.e., servers) at each off-link. Suppose that there are m off-ramps at a sanitary cordon. It is necessary to study the entire queuing network performances. Assume that vehicles arrive at i-th (i=1,2,…,m) off-ramp according to a Poisson process with predicted inflow λ_i and that i-th off-ramp has an exponential service time distribution with an identical parameter μ for its c_i parallel checkpoints, where $c_i\mu>\lambda_i$. Therefore, the elementary $M/M/c$ queuing model can be used to analyse each off-ramp.

Analogous to a single service facility, the most commonly used measures of queuing situation in a given off-ramp i are the expected number of vehicles (l_i) and expected delay time (d_i). Note that l_i includes the expected number of vehicles in the queue and in service, and d_i includes expected waiting time in the queue and expected service time. The relationship between l_i and d_i is known as Little's formula, and it is given as $l_i=\lambda_i d_i$. The relationship is valid under rather general conditions. Let $\rho_i=\lambda_i/\mu$, the expression l_i can be determined as follows:

$$l_i = \frac{\rho_i^{c_i+1}}{(c_i-1)!(c_i-\rho_i)^2} p_i + \rho_i, \quad \forall i \in A^* \tag{1}$$

$$\rho_i = \left[\sum_{n=0}^{c_i-1} \frac{\rho_i^n}{n!} + \frac{\rho_i^{c_i}}{(c_i-1)!(c_i-\rho_i)}\right]^{-1}, \quad \forall i \in A^* \tag{2}$$

$$\frac{\rho_i}{c_i} < 1, \quad \forall i \in A^* \tag{3}$$

where p_i is the steady-state probability of none customers in an off-ramp i, A^* is the set of off-links, *Equation 3* is a steady-state condition. The measure d_i is determined through dividing l_i by λ_i according to Little's formula. It can be formulated in detail as follows:

$$d_i(\lambda_i, c_i) = \frac{1}{\lambda_i}\left[\frac{\rho_i^{c_i+1}}{(c_i-1)!(c_i-\rho_i)^2}\left[\sum_{n=0}^{c_i-1} \frac{\rho_i^n}{n!} + \frac{\rho_i^{c_i}}{(c_i-1)!(c_i-\rho_i)}\right]^{-1} + \rho_i\right], \quad \forall i \in A^* \tag{4}$$

The travel times at off-links consist of two parts. One is link travel time determined by link traffic flows. The other is queuing delay time determined by traffic flow and the number of parallel checkpoints. Note that the queues are assumed to have zero physical length, and there is no queue spillback, which implies that link travel time is independent of the length of the queues. Let the generalised travel time be defined as $t_a(v_a,c_a)=t_a(v_a)+d_a(v_a,c_a)$, if $a\in A^*$ and $t_a(v_a,c_a)=t_a(v_a)$ otherwise where $t_a(v_a)$ is link travel time and $d_a(v_a,c_a)$ is queuing delay time. The traffic flows v_a at an off-link is regarded as the arrival rate of the queuing system. Then the traffic assignment with queuing for a metering scheme **u** is a conventional problem.

In summary, the lower-level model is conducted as follows, given travel demand and road network. Trip distribution is generated by aggregating individual destination choice. The multinomial logit model is used for destination choices, and it is regarded as the simplest and most practical. After the trip distribution matrix is generated, travel demands are assigned into the road network by user equilibrium with queuing to generate link traffic flows. Note that the predicted traffic flow at each link is regarded as the average arrival rate of the queuing system. Then all the origin-destination (O-D) pair travel times are produced by Dijkstra's algorithm with link travel times and queuing delay times. These path travel times are fed back to the multinomial logit model to update the trip distribution matrix. This process is iterated until the trip distribution matrix is well-converged. The state is known as transport system equilibrium. *Figure 2* illustrates the feedback process at the lower-level.

Variable notations used in *Figure 2* are defined as follows:

- q_{rs} – travel demand between origin r and destination s;
- u_r – traffic inflow through the ramp at origin r;
- s_r – set of destinations for travellers departed from origin r;
- β_s – traveller preference for destination s;
- t_{rs} – path travel time between origin r and destination s;
- A – set of links in the network;
- A^* – set of off-links that is a subset of links in the network;
- β_t – coefficient of travel time t_{rs};
- v_a – traffic flow at the link a;
- c_a – number of checkpoints at the link a;
- t_a – travel time at the link a which is a function of traffic flow v_a and number of checkpoints c_a;
- f_k^{rs} – traffic flow on the path k connecting origin r and destination s;
- $\delta_{a,k}^{rs}$ – link-path incidence relationship which is expressed as:

$$\delta_{a,k}^{rs} = \begin{cases} 1, \text{if link } a \text{ is on path } k \text{ connecting } r \text{ and } s \\ 0, \text{if not} \end{cases}.$$

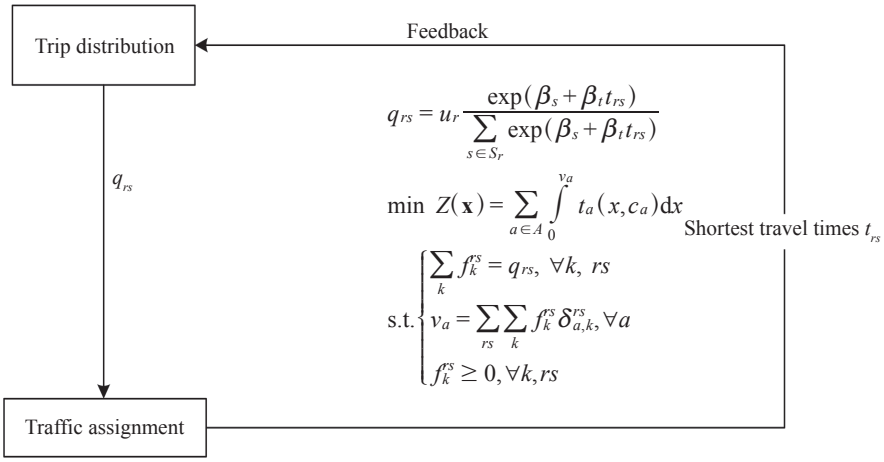

Figure 2 – The iterative process in the lower-level model

2.2 The bi-level programming model

There is a transportation system equilibrium with queuing in the lower level for a given on-ramp metering scheme in the upper level. The operator aims to maximise network throughput with a predetermined level of waiting time at each off-ramp by metering on-ramps. Acceptable control should consider queuing delay constraints. The aspiration level model works directly with the performance measures of the queuing situation. The idea is to determine an acceptable range for the service level by specifying reasonable limits on measures of performance. Such limits are the aspiration levels the decision-maker wishes to reach. Note that the service level in a given off-ramp i is a function of the link flow λ_i which could be controlled by on-ramp meters. This step presents a constraint for determining acceptable on-ramp metering rate considering the average waiting time such as

$$d_i(\lambda_i, c_i) \leq T,\ \ \forall i \tag{5}$$

The constant T in *Equation 5* is the level of aspiration specified by the decision-maker, for example, T=3 minutes. Note that d_i is a function of link flow λ_i and number of checkpoints c_i. According to *Equation 4*, the constraint on average waiting time d_i can be specified in more detail as follows:

$$\frac{1}{\lambda_i}\left[\frac{\rho_i^{c_i+1}}{(c_i-1)!(c_i-\rho_i)^2}\left[\sum_{n=0}^{c_i-1}\frac{\rho_i^n}{n!}+\frac{\rho_i^{c_i}}{(c_i-1)!(c_i-\rho_i)}\right]^{-1}+\rho_i\right] \leq T \tag{6}$$
$$\forall i \in A^*$$

A bi-level optimisation method was proposed by Yang et al. [17], where the on-ramp traffic control problem was described as a leader-follower or a Stackelberg game. With the inspiration of their work, a bi-level programming model is also adopted here. The freeway operator in the upper-level aims to maximise the total input flow through all ramps with queuing delay constraint. The operator can predict, but cannot control travel behaviour of freeway users including destination choice and route choice, while all users make their decisions in a user optimal manner. The upper-level model is formulated as follows:

$$\text{Max}\ U = \sum_{r \in R} u_r \tag{7}$$

subject to

$$\frac{\rho_i^{c_i+1}}{\lambda_i(\mathbf{u})(c_i-1)!(c_i-\rho_i)^2}\left[\sum_{n=0}^{c_i-1}\frac{\rho_i^n}{n!}+\frac{\rho_i^{c_i}}{(c_i-1)!(c_i-\rho_i)}\right]^{-1} \leq T \tag{8}$$
$$i \in I$$

$$\frac{\lambda_i(\mathbf{u})}{c_i\mu} < 1,\ \ \forall i \in I \tag{9}$$

$$0 \leq u_r \leq u_r^d,\ \ r \in R \tag{10}$$

where

u_r – inflow through on-ramp $r \in R$;
u_r^d – traffic demand at on-ramp $r \in R$;
R – set of on-ramps, i.e., origins;
I – set of off-ramps;
$\mathbf{u}$ – vector of all inflows (upper-level decision variables).

In the upper-level model, *Equation 7* is the objective function to maximise the total input flow through all on-ramps, *Equation 8* means that the waiting time at each off-ramp should not exceed a predetermined acceptable level T. Note that the link flow $\lambda_i(\mathbf{u})$ is formulated as a function of all inflows, *Equation 9* is the steady-state condition, and *Equation 10* indicates that the inflow through each ramp should be non-negative and equal to or less than the corresponding traffic demand. The link flow $\lambda_i(\mathbf{u})$ is obtained by solving the lower-level problem.

3. SOLUTION ALGORITHM

3.1 Equilibrium algorithm with queuing

To solve the proposed bi-level programming model, it is always beneficial to solve the lower-level model first as it is embedded in the upper-level model. With a built road network, there will be a stable flow pattern in the lower-level for a given ramp metering scheme from the upper-level. Note that the lower-level is a feedback procedure between a trip distribution and traffic assignment with queuing. The method of successive averages (MSA) can be used to achieve system equilibrium. An initial trip distribution matrix can be produced by a multinomial logit model with initialised origin-destination (O-D) pair travel times. The trips are then assigned the road network according to the user equilibrium principle. The link travel flows and link travel times can be generated by the Frank-Wolfe algorithm [10]. In addition, queuing delay times can be determined with predicted traffic flows at off-links. The generalised travel time of each off-link includes link travel time and queuing delay time. According to Wardrop's first principle of route choice, also known as user equilibrium, traffic arranges itself in congested networks such that all used paths between an O-D pair have an equal and minimum cost. Therefore, Dijkstra's algorithm is used to update the O-D pair travel times. These times are then fed back to the multinomial logit model to generate a new trip distribution matrix. However, this matrix cannot be assigned to the road network directly. The convergence of direct or naive feedback is usually impossible. An averaging of successive trip distribution matrix is necessary. Although there are some successful applications of constant weights, the convergence is usually not guaranteed. Therefore, the MSA with decreasing weight is used here to update the trip distribution matrix, which is the reciprocal of the iteration number. The updated matrix is further assigned to the road network. The iteration process continues until the successive matrices are quasi-equal. The convergence is generally measured by the squared root of the relative gap between successive travel demand matrices. If a predetermined tolerance is achieved, terminate the iteration. The stable state is known as the transportation system equilibrium. The resultant traffic inflows at all the off-links then go into the upper-level model. *Figure 3* shows the flowchart of the equilibrium algorithm with queuing.

The detailed MSA algorithm is specified step by step as follows:

Step 1: Input an on-ramp metering scheme decision **u** and a built road network.

Step 2: Initialise trip distribution matrix q_{rs}^0 with initial O-D pair travel time t_{rs}^0. Besides, let n=1 be the number of iterations.

Step 3: Traffic assignment with queuing. The trip distribution matrix q_{rs}^0 is assigned to the road network by the Frank-Wolfe algorithm [10]. The link travel flows v_a and link travel times t_a are generated. Note that queuing delay times can be determined with flows at off-links.

Step 4: Update the shortest path travel time between an O-D pair rs, namely t_{rs}^1, by Dijkstra's algorithm.

Step 5: Trip distribution. The multinomial logit model is used to update the trip distribution matrix q_{rs}^1:

$$q_{rs}^1 = u_r \frac{\exp(\beta_s + \beta_t t_{rs}^1)}{\sum_{s \in Sr} \exp(\beta_s + \beta_t t_{rs}^1)} \tag{11}$$

Step 6: Average trip distribution matrices q_{rs}^1 and q_{rs}^0 using decreasing weight

$$q_{rs}^1 = q_{rs}^0 + \frac{1}{n}(q_{rs}^1 - q_{rs}^0) \tag{12}$$

Step 7: Convergence identification. Check the convergence of trip distribution matrix using the squared root of the relative gap:

$$\sqrt{\sum_{rs}\left(\frac{q_{rs}^1 - q_{rs}^0}{q_{rs}^0}\right)^2} < \varepsilon \tag{13}$$

where ε is a predetermined tolerance. If the convergence condition is satisfied, terminate the iteration and go to Step 9, otherwise go to Step 8.

Step 8: Let $q_{rs}^0 := q_{rs}^1$ and $n := n+1$. Then go to Step 3.

Step 9: The outputs are the trip distribution matrix q_{rs}^1 and the link traffic flow v_a.

3.2 The solution of the bi-level programming problem

The bi-level programming problem is a well-known NP-hard problem that is difficult to solve by classical optimisation algorithms. It is challenging even if the upper-level and lower-level are both linear programming models, let alone the upper-level is a nonlinear programming model. For example, the traditional gradient-based approaches to solve the optimal cordon toll problem usually fail to converge for larger-scale problems due to multiple optima. This failure led to the development of a heuristic algorithm to determine the optimal toll level

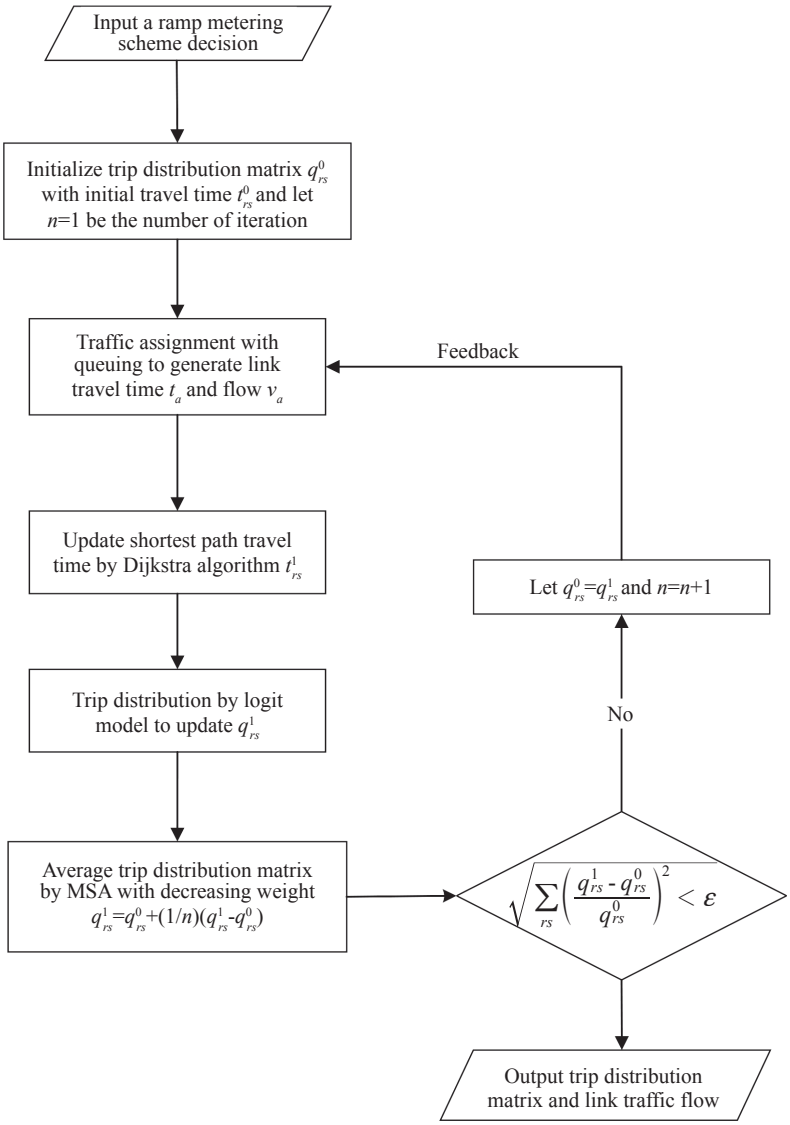

Figure 3 – The flowchart of the MSA algorithm

and toll location problem. The heuristic algorithm was shown to be successful in solving the cordon toll optimisation problem, although it is found to be time-consuming, and there is no proof of global optimum. However, the successful applications of heuristic methods, especially genetic algorithms, have been growing to generate high-quality cordon schemes in the literature [18–21]. Therefore, a genetic algorithm with an elite strategy is adopted here. *Figure 4* shows its flowchart.

To be more specific, the detailed genetic algorithm with an elite strategy is specified in steps as follows:

Step 1: Initialisation. Set the parameters used in the genetic algorithm, including population size M, the maximum number of generations Gen, crossover probability p_c, mutation probability p_m, the notation of generation gen=1, the portion for elitist strategy p_e. Note that the population size depends on the nature of the problem, but typically contains several hundreds of possible solutions.

Step 2: Generate a feasible initial population randomly. A chromosome is a solution that consists of several genes. Note that the number of genes in a chromosome is equal to the number of on-ramp meters which are decision variables in the upper-level. Real encoding technology is used where a gene stands for the on-ramp metering rate at an off-link. Generate a chromosome randomly. If it is not feasible, generate another one until it is feasible. A total number of M viable chromosomes are generated, scattering the entire range of possible solutions.

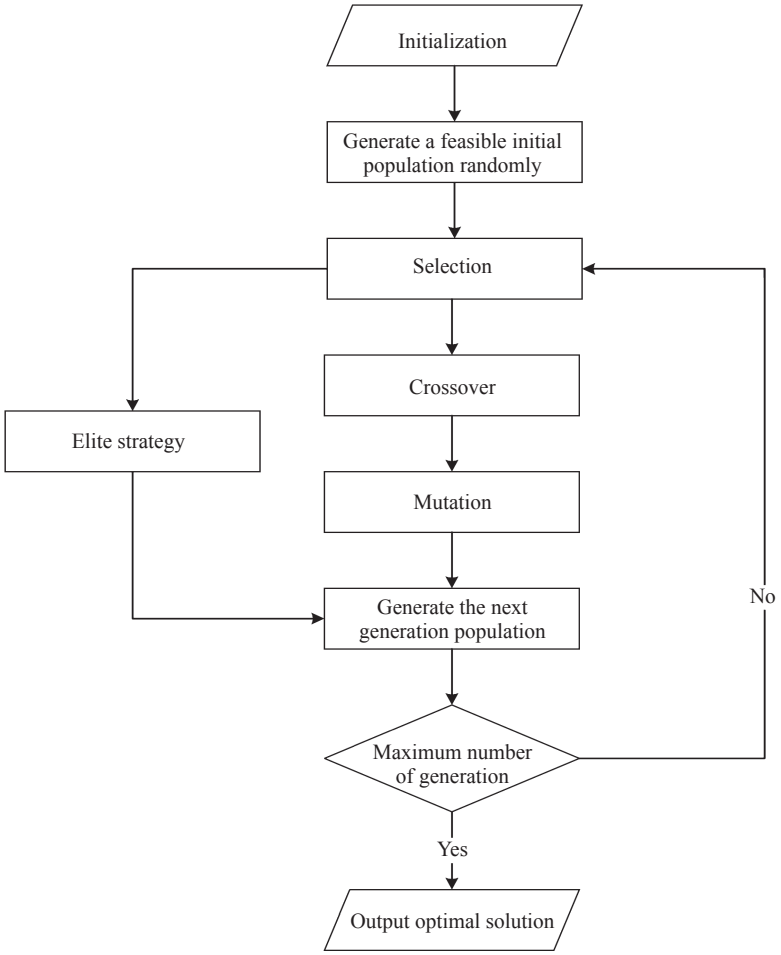

Figure 4 – The flowchart of a genetic algorithm with the elite strategy

Step 3: Selection operation. The objective function of the upper-level model is used to work as a fitness function to evaluate the performance of all chromosomes in the population. Note that to maximise the network throughput, the best p_e is labelled for elitists and the worst p_e is discarded.

Step 4: Crossover operation. The remaining $(1\text{-}p_e)M$ chromosomes are used for crossover operation. These parent chromosomes are matched in pairs randomly. The probability of carrying out the crossover is p_c. If it is chosen for the crossover, a random gene is identified. If new-born chromosomes are not feasible according to constraints in the upper-level model, try another gene location until they are feasible. These new solutions typically share many of the characteristics of their parents.

Step 5: Mutation operation. The probability of carrying out mutation is p_m. A random gene is identified for mutation within the domain of definition. If the new chromosome is not feasible, try another gene location until it is feasible.

Step 6: Generate the next generation population. After genetic operators, there are still $(1\text{-}p_e)M$ feasible chromosomes. The labelled p_eM elitists are added to ensure the population size M. This allows the best chromosomes from the current generation to carry over the next unaltered. It guarantees that the solution quality will not decrease from one generation to the next. Let the notation of generation be $gen{:=}gen{+}1$.

Step 7: Termination judgment. If the maximum number of generations is achieved, that is $gen \geq Gen$, terminate the iteration process and output the optimal scheme of on-ramp metering. Otherwise, go to Step 3.

4. EXPERIMENTAL STUDY

In order to verify the effectiveness of the proposed method and algorithm, an experimental study is conducted. The Nguyen-Dupuis road network, as shown in *Figure 5* is commonly used in transportation research to demonstrate various methods. The

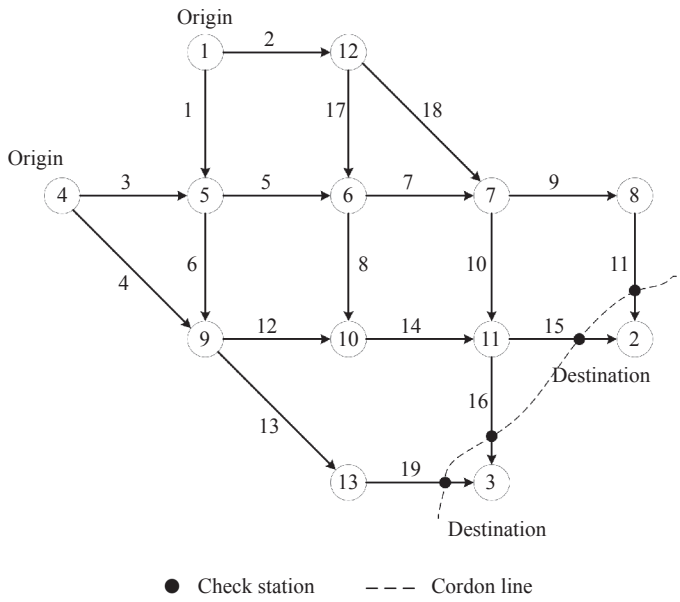

Figure 5 – The Nguyen-Dupuis road network

link characteristics, including free-flow travel time, link capacity, and link length, are shown in *Table 1*. Note that it is a closed road network with two origins being controlled as the entry points and two exits containing a sanitary cordon. The background stream outside the Nguyen-Dupuis road network is not considered at present.

There are two origins and two destinations in the Nguyen-Dupuis network. The maximum travel demands at origin nodes 1 and 4 are 1,500 veh/h and 1,500 veh/h respectively. That is, u_1^d=1,500 veh/h and u_2^d=1,500 veh/h. There is a ramp meter at each origin to control travel demand going into the freeway network. All of the existing links are labelled from 1 to 19. The destination nodes are 2 and 3 so the off-links to destinations are 11, 15, 16, 19. The problem is to determine the on-ramp metering rate at each origin in order to maximise the total input flow with an acceptable waiting time at cordon sanitaire.

The parameters used in the lower-level model are summarised as follows. The multinomial logit model for destination choices is simplified as

$$q_{rs}^1 = u_r \frac{\exp\left(\beta_s + \beta_t t_{rs}^1\right)}{\sum_{s \in Sr} \exp\left(\beta_s + \beta_t t_{rs}^1\right)} \tag{14}$$

where β_s is traveller preference on destination s and β_t is the coefficient of path travel time between O-D pair rs. The values of β_s and β_t can be calibrated empirically. Here we set β_2=0.5, β_3=0, and β_t=-0.1. That is, the traveller preference on destination node

Table 1 – Link characteristics of the Nguyen-Dupuis road network

Link *a*	Free flow time [min]	Link capacity [veh/h]	Link length [km]
1	7.0	800	4.00
2	9.0	800	6.00
3	9.0	800	5.00
4	12.0	800	8.00
5	3.0	800	2.00
6	9.0	800	5.00
7	5.0	800	3.00
8	13.0	800	8.00
9	5.0	800	3.00
10	9.0	800	6.00
11	9.0	800	5.00
12	10.0	800	6.00
13	9.0	800	5.00
14	6.0	800	4.00
15	9.0	800	6.00
16	8.0	800	5.00
17	7.0	800	4.00
18	14.0	800	6.00
19	11.0	800	7.00

2 is 0.5, and on destination node 3 is 0 which means that the travellers traditionally prefer destination 2. The coefficient of travel time is -0.1, which means that the travel time is a negative utility. In addition, a link impedance function, named BPR function, is used to accommodate congestion effect in traffic assignment with the following formulation:

$$t_a(v_a) = t_a^0\left[1+\alpha\left(\frac{v_a}{e_a}\right)^\beta\right],\ a \in A \tag{15}$$

where t_a^0 is the free-flow travel time of link a; α and β are volume/delay coefficients which can be calibrated empirically, they are usually set as α=0.15 and β=4 conventionally; e_a is road capacity of link a. The other notations are consistent with previous definitions. The convergence criteria for MSA is set as ε=0.01. A stable transportation system can be achieved for a built road network.

The lower-level model is embedded in the upper-level model. The parameters used in the upper-level model are listed as follows. The population size is M=500. The maximum number of generations is Gen=30. The portion for elitists is p_e=0.1. The crossover probability is p_c=0.1, and the mutation probability is p_m=0.5. Although these parameters are conventionally used in genetic algorithms, it is worth tuning parameters to find reasonable settings for the problem. The maximum acceptable waiting time T at each off-ramp is assumed to be 2 minutes. The number of testing checkpoints is assumed to be 9, 3, 5, 5 for off-links 11, 15, 16, 19, respectively according to their physical conditions.

Note that traffic volumes are usually measured in hours (veh/h) while the arrival rates are usually measured in minutes (veh/min). Therefore, unit conversion is needed. The average service rate for a single checkpoint is assumed to be μ=2 veh/min. That is the checkpoint averagely tests two vehicles per minute. The calculation is programmed using the popular open-source language R 3.6.3 in a personal computer with Intel Core i7-4790 CPU @ 3.60GHz. The running time is 2.8 hours. The maximum network throughput is 2,366 veh/h. The on-ramp metering rate at origin 1 is 1,385 veh/h, and that at origin 2 is 981 veh/h. The network performances for all links with checkpoints are shown in *Table 2*.

It is shown that traffic inflow volumes are different at off-links. The maximum one is 1,025 veh/h at off-link 11 while the minimum one is only 319 veh/h at off-link 15. The waiting time at each off-link is different, but they do not exceed the predetermined waiting time. The maximum waiting time is 1.660 min at off-link 15, and the minimum waiting time is 0.866 min at off-link 19. Although the GA based algorithm can give an approximated optimal solution, attention should be paid to the issue of choosing the best parameters, i.e., generation number, population number, probability of crossover, and the probability of mutation.

Table 2 – The queuing performance with maximum network throughput

Off-link	Traffic inflow volume [veh/h]	Number of checkpoints	Waiting time [min]
11	1,025	9	1.407
15	319	3	1.660
16	522	5	1.036
19	499	5	0.866

5. CONCLUSION

This paper proposed a method to determine the optimal on-ramp metering rates for maximum network throughput with a predetermined queuing delay constraint at off-ramps around cordon sanitaire. A bi-level programming model is formulated where the upper-level is on-ramp traffic control problem, and the lower-level is transportation system equilibrium problem. To be more specific, the lower-level is a closed-loop feedback procedure between trip distribution and traffic assignment given ramp control and freeway network. The iteration continues until the transportation system equilibrium is achieved. Note that a multinomial logit model is used for trip distribution and user equilibrium model is used for traffic assignment. The traffic flow at each off-link can be predicted in this way. The upper-level is to design an on-ramp traffic control scheme with an acceptable queuing delay constraint at cordon sanitaire. The traffic flow at each off-ramp is regarded as an average arrival rate in queuing model. Note that there could be many off-ramps and the $M/M/c$ queuing model is used to analyse each off-ramp independently of the others. A nonlinear programming model is built in the upper level where the objective is to maximise the total input flow through all on-ramps.

A heuristic algorithm is proposed to solve the bi-level optimisation model. Method of successive averages (MSA) is used to achieve transportation system equilibrium at the lower-level model. Note that the conventional Frank-Wolfe algorithm is used

for traffic assignment. A genetic algorithm with elite strategy is designed to solve the nonlinear programming model at the upper-level. As the lower-level model is embedded in the upper-level model, there is a transportation system equilibrium for a given feasible on-ramp control scheme.

To demonstrate the effectiveness of the proposed method and algorithm, an experimental study is conducted using the well-known Nguyen-Dupuis road network. Although the test is only limited to a simplified network, the results show that the methods can find at least a good heuristic optimal solution. The maximum network throughput can be determined. However, it is worth tuning the parameters to find reasonable settings for the problem.

There are some recommendations for future research. Firstly, a large-scale network with real data will be more convincing while this paper only adopted a simplified network for demonstration. Otherwise, the suggestions for policy makers will be more valuable. Secondly, the deployment of checkpoints could be jointly determined with on-ramp metering while it is assumed to be fixed here. However, it is no wonder that this will make the model and computation much tougher. Finally, more efficient algorithm is needed for global optimisation. Although genetic algorithm is accessible, it is time-consuming.

ACKNOWLEDGMENT

The authors would like to acknowledge the extremely useful comments offered by the anonymous referees. This research is supported by the natural science foundation of Jiangsu province (Grants No. BK20201277) and the major project of philosophy and social science research in colleges and universities in Jiangsu province (Grants No. 2021SJZDA025).

林宏志，博士
电子邮箱：linhz@seu.edu.cn
东南大学经济管理学院
南京市江宁区东南大学路2号

用于新型冠状病毒防控的城市高速公路网进口匝道优化控制方法

摘要：

新型冠状病毒肺炎（*COVID-19*）的爆发扰乱了人们的正常生活。地方政府普遍采用了设置防疫封锁线的措施来构建城市安全屏障，出行的人们只有经过检测后才能通过封锁线。本文旨在提出一种进口匝道优化控制方法，以防疫封锁线上安全检查站处的排队延迟时间为约束，使城市高速公路网络的通行能力最大化。为此建立了双层规划模型，其中下层是预测道路流量的交通系统平衡模型，上层是用于优化控制的非线性规划模型。尤其是采用了随机排队模型描述在安全检查站处的等待现象，并设计了启发式算法求解该双层模型，其中下层模型采用连续平均算法（*MSA*），上层模型采用具有精英策略的遗传算法（*GA*）。最后，进行了实验研究以验证所提出的方法和算法的有效性。结果表明，该方法可以找到一个良好的启发式最优解。这些方法对于高速公路运营者确定用于疾病防控的最优匝道控制方法十分有效。

关键词：防疫封锁线；进口匝道控制；双层规划模型；启发式算法；排队论

REFERENCES

[1] Quilty BJ, et al. The effect of travel restrictions on the geographical spread of COVID-19 between large cities in China: A modelling study. *BMC Medicine*. 2020;18(1): 259. doi: 10.1101/2020.02.09.20021261.

[2] Gostin LO, Wiley LF. Governmental public health powers during the COVID-19 pandemic: Stay-at-home orders, business closures, and travel restrictions. *JAMA - Journal of the American Medical Association.* 2020;323(21): 2137-8. doi: 10.1001/jama.2020.5460.

[3] Pan A, et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *JAMA - Journal of the American Medical Association.* 2020;323(19): 1915-23. doi: 10.1001/jama.2020.6130.

[4] Yang H, Yagar S. Traffic assignment and traffic control in general freeway-arterial corridor systems. *Transportation Research Part B: Methodological*. 1994;28(6): 463-86. doi: 10.1016/0191-2615(94)90015-9.

[5] Yang H, Yagar S. Traffic assignment and signal control in saturated road networks. *Transportation Research Part A: Policy and Practice.* 1995;29(2): 125-39. doi: 10.1016/0965-8564(94)E0007-V.

[6] Yang H, Lam WHK. Optimal road tolls under conditions of queueing and congestion. *Transportation Research Part A: Policy and Practice*. 1996;30(5 PART A): 319-32. doi: 10.1016/0965-8564(96)00003-1.

[7] Yang H, Bell MGH. Traffic restraint, road pricing and network equilibrium. *Transportation Research Part B: Methodological*. 1997;31(4): 303-14. doi: 10.1016/S0191-2615(96)00030-6.

[8] Vickrey WS. Congestion theory and transport investment. *The American Economic Review.* 1969;59(2): 251-60. http://www.jstor.org/stable/1823678 [Accessed 27th Apr. 2021].

[9] Li Z-C, Huang H-J, Yang H. Fifty years of the bottleneck model: A bibliometric review and future research directions. *Transportation Research Part B: Methodological.* 2020;139: 311-42. doi: 10.1016/j.trb.2020.06.009.

[10] Sheffi Y. Urban transportation networks. Englewood Cliffs, NJ: Prentice-Hall; 1985.

[11] Oppenheim N. Urban travel demand modeling: from individual choices to general equilibrium. New York: John Wiley and Sons; 1995.

[12] Boyce DE, Zhang Y-F, Lupa MR. Introducing "feedback" into four-step travel forecasting procedure versus equilibrium solution of combined model. *Transportation Research Record.* 1994;1443: 65-74. http://worldcat.org/isbn/0309055245 [Accessed 27th Apr. 2021].

[13] Boyce D, Zhang Y-F. Calibrating combined model of trip distribution, modal split, and traffic assignment. *Transportation Research Record.* 1997;1607: 1-5. doi: 10.3141/1607-01.

[14] Lin H. An accessibility-oriented optimal control method for land use development. *Journal of Urban Planning and Development*. 2019;145(4): 04019011. doi: 10.1061/(ASCE)UP.1943-5444.0000518.

[15] Lin H-Z, Wei J. Optimal transport network design for both traffic safety and risk equity considerations. *Journal of Cleaner Production*. 2019;218: 738-45. doi: 10.1177/0037549720920374.

[16] Gartner NH, Messer CJ, Rathi A. Revised Monograph on Traffic Flow Theory. Federal Highway Administration, United States; 1999.

[17] Yang H, et al. An algorithm for the inflow control problem on urban freeway networks with user-optimal flows. *Transportation Research Part B: Methodological*. 1994;28(2): 123-39. doi: 10.1016/0191-2615(94)90021-3.

[18] Shepherd S, Sumalee A. A genetic algorithm based approach to optimal toll level and location problems. *Networks and Spatial Economics*. 2004;4(2): 161-79. doi: 10.1023/B:NETS.0000027771.13826.3a.

[19] Sumalee A. Optimal road user charging cordon design: A heuristic optimization approach. *Computer-Aided Civil and Infrastructure Engineering*. 2004;19(5): 377-92. doi: 10.1111/j.1467-8667.2004.00364.x.

[20] Sumalee A. Multi-concentric optimal charging cordon design. *Transportmetrica*. 2007;3(1): 41-71. doi: 10.1080/18128600708685667.

[21] Liu Z, Meng Q, Wang S. Speed-based toll design for cordon-based congestion pricing scheme. *Transportation Research Part C: Emerging Technologies*. 2013;31: 83-98. doi: 10.1016/j.trc.2013.02.012.