TOWARD A GENERAL FRAMEWORK FOR DYNAMIC ROAD PRICING

Andy H.F. Chow, Centre for Transport Studies, University College London.

ABSTRACT

This paper develops a general framework for analysing and calculating dynamic road toll. The optimal network flow is first determined by solving an optimal control problem with state-dependent responses such that the overall benefit of the network system is maximized. An optimal toll is then sought to decentralise this optimal flow. This control theoretic formulation can work with general travel time models and cost functions. Deterministic queue is predominantly used in dynamic network models. The analysis in this paper is more general and is applied to calculate the optimal flow and toll for Friesz’s whole link traffic model. Numerical examples are provided for illustration and discussion. Finally, some concluding remarks are given.

1. INTRODUCTION

To capture the transient nature of traffic congestion and the so-called “peak spreading” effect (Small, 1992), dynamic network models have been developed in which the traffic flows in a network and the consequent travel costs are considered to be varying over time. A dynamic network model comprises three interacting components: a network loading model, an elastic travel demand function, and a traffic assignment model. The network loading model captures the propagation of traffic and determines the costs of travel. The travel demand function specifies the amount of traffic generated between each origin-destination pair in the network within a fixed time horizon according to the travel costs. The assignment model determines the network flows given the travel costs that the travellers encounter. We consider two assignment principles: dynamic equilibrium and system optimal assignments. In equilibrium, the total travel costs experienced by travellers are equal and minimal for each origin-destination in the network. In system optimal, travellers are assigned such that the total
Mathematics in Transport Planning and Control

travellers’ surplus in the network is maximized. The total costs incurred differ between travellers at system optimal. A dynamic toll is then sought to complement the private costs incurred by the travellers so that the system converts back to an equilibrium state.

The seminal work on modelling and managing dynamic traffic was by Vickrey (1969). This paper extends Vickrey’s model to a general framework for analysing and calculating optimal time-varying toll. In the next section, we start with reviewing various network loading models and travel cost functions. In section three, we have a brief discussion on dynamic equilibrium assignment. To derive the optimal toll, we need to formulate a system optimization problem. In section four, we formulate this as an optimal control problem with state-dependent response. We derive the optimality conditions for this special kind of control problem using calculus of variations. At optimality, traffic is assigned such that the total travellers’ surplus in the network is maximized. It is then solved by a dynamic-programme solution algorithm. It should be noted that the total costs incurred differ between travellers at optimality and hence the system is not in equilibrium. A dynamic toll is then sought to complement the private costs incurred by the travellers so that the system converts to an equilibrium state. Numerical examples are presented for illustration in section five. Finally, concluding remarks are given in section six.

2. NETWORK LOADING MODELS AND COST FUNCTIONS

The network loading model determines the corresponding time-varying flows and travel times given the network inflow. The model is considered to be plausible if it satisfies: positivity of flows; First-in-first-out (FIFO) principle; flow conservation principle; flow propagation principle and causality. Detailed discussions on these are referred to Carey (2004). To ensure FIFO, Daganzo (1995) concluded that the model should only depend on the traffic on link. The remaining possibilities can then be divided into two categories: outflow models and travel time model. However, outflow models have been extensively criticized for their implausible traffic propagation behaviour and violation of causality (Astarita, 1996; Heydecker and Addison, 1998). Astarita (1996) and Mun (2002) and others further demonstrated that FIFO cannot be guaranteed if the travel time model is non-linear in link. Consequently, this paper only considers linear travel time models.

2.1 A class of linear travel time models

We consider that each link $a$, which has a flow-invariant travel time $\phi_a$ and a link capacity $Q_a$, comprises two parts as shown in Figure 1. The portion $\alpha_a$ represents the “congestible” part of the link and hence $\phi_a - \alpha_a$ is the “free flow” part. A general form for this class of linear link travel time models can be written as:

$$\tau_a(s) = s + \phi_a + x_a(s + \phi_a - \alpha_a)/Q_a,$$  \hspace{1cm} (1)
where $s$ represents the time of entry to the link and $\tau_a(s)$ is the corresponding time of exit. The amount of link traffic in the congestible part is represented by $x_a(s)$.

![Figure 1: Representation of a travel link](image)

**2.1.1 Deterministic queuing model**

Vickrey (1969) considered each link corresponds to a freely flowing link with a flow-invariant travel time $\phi_a$ (i.e. $\alpha_a = 0$) with a deterministic queue at its downstream end with a maximum service rate $Q_a$. We name this travel time model as “deterministic queuing” model. The state equation of $x_a(s)$ of this model is given by

$$\frac{dx_a(s)}{ds} = e_a(s - \phi_a) - g_a(s).$$

Whenever a queue exists, the link outflow is equal to the capacity and all travellers arrive before the queue dissipates will incur travel delay. Otherwise, if the queue length is zero, the outflow is taken as the inflow at the time of entry and the travellers are unimpeded. i.e.:

$$g_a(s) = \begin{cases} e_a(s - \phi_a) & (x_a(s) = 0, e_a(s - \phi_a) < Q_a) \\ Q_a & \text{otherwise} \end{cases}.$$

**2.1.2 Whole-link traffic model**

Friesz et al. (1993) proposed a linear travel time model that considers the whole travel link to be congestible, i.e. $\alpha_a = \phi_a$. We regard this travel time model as “whole-link” traffic model. The state equation of $x_a(s)$ of this model is given by

$$\frac{dx_a(s)}{ds} = e_a(s) - g_a(s).$$

The outflow experienced by traffic that enters at time $s$ can be established according to correct flow propagation (Heydecker and Addison, 1998) as

$$g_a[\tau(s)] = \frac{Q_a e_a(s)}{Q_a + e_a(s) - g_a(s)},$$

which depends on outflows at time $s$ and hence on inflows at earlier times.
Mathematics in Transport Planning and Control

2.1.3 Divided linear travel time model

Mun (2002) proposed another linear travel time called “divided linear travel time” model by letting \( \alpha \) be the size of the time incremental step \( \Delta s \) in discretization. The model has been shown to be able to give plausible traffic behaviour. While, a detailed discussion of this travel time model is beyond the scope of the present paper.

2.2 Numerical comparisons of the travel time models

Although the functional forms of the two travel time models look very much similar, they behave quite differently. Figure 2 compares the travel times and the outflow profiles calculated by these two models. We load a parabolic inflow into a travel link with free flow travel time \( \phi \) equals to 2mins and capacity \( Q \) equals to 30 veh/min. Contrast with the deterministic queue, the outflow varies continuously with the inflow over time for the whole-link traffic model. The outflow will approach to, but not exceed, the link capacity for a high inflow rate. Moreover, the travel time estimated by whole-link traffic model is substantially higher as the whole-link model considers the whole travel link to be congestible.

![Figure 2 Numerical comparisons of the travel time models](image)

2.3 Travel cost functions

We suppose that travel behaviour can be represented as a response to the various costs associated with travel. We consider the travel cost encountered by each traveller has three distinct components. The first component is the travel time which is determined by the travel time model as discussed previously. In addition to the travel time, we add a time-specific cost \( f(\tau_p(s)) \) associated with arrival time \( \tau_p(s) \) at the destination. Finally, we add a time-specific cost \( h(s) \) associated with departure from the origin at time \( s \). Possible choices of these time-specific cost functions are investigated by Heydecker and Addison (2005). Following these specifications, the total travel cost \( C_p(s) \) associated with departure on route \( p \) at time \( s \) is defined as a linear combination of these costs as
\[ C_p(s) = h(s) + [\tau_p(s) - s] + f[\tau_p(s)]. \]  

(6)

3. DYNAMIC EQUILIBRIUM CONDITION

Hendrickson and Kocur (1981) showed that if departure time choice is considered together with route choice, then the total travel cost \( C_p(s) \) incurred will take a single value for each origin-destination pair in the network in equilibrium. This can be stated as a complementary inequality for the inflow \( e_p(s) \):

\[
e_p(s) \begin{cases} > 0 \Rightarrow C_p(s) = D_{od}^{-1}[E_{od}(T)] \\ = 0 \Rightarrow C_p(s) \geq D_{od}^{-1}[E_{od}(T)] \end{cases}, \forall p \in P_{od}, \forall od, \forall s,
\]

(7)

where \( P_{od} \) is the set of all routes from \( o \) to \( d \) and \( D_{od}^{-1}[E_{od}(T)] \) is the total cost at which travel takes place from \( o \) to \( d \), given the total throughput \( E_{od}(T) \). Using the first case in (7) and differentiating both sides with respect to the departure time \( s \), we have:

\[
e_p(s) > 0 \Rightarrow \Omega_p(s) = h'(s) + \tau_p(s) - 1 + f'[\tau_p(s)]\tau_p(s) = 0.
\]

(8)

Hence, solving the equilibrium assignment is equivalent to solving the set of simultaneous equations \( \Omega_p(s) = 0 \) for all routes \( p \) in use.

4. SYSTEM OPTIMAL ASSIGNMENT AND EXTERNALITY

The system optimal assignment seeks an optimal inflow \( e_p^*(s) \) that maximizes the total travellers’ surplus in the network within a fixed planning period \( T \). The assignment is formulated as the optimal control problem:

\[
\max_{e_p(s)} Z = \sum_{\forall od} \int_0^T D_{od}^{-1}(\omega) d\omega - \sum_{\forall od} \sum_{\forall p} \int_0^T C_p(s)e_p(s) ds
\]

(9a)

subject to:

\[
e^p_{a_m}(\tau^p_{a_m}(s)) = \frac{e^p_{a_m}(s)}{\tau^p_{a_m}(s)}, \forall a_m \in A_p, \forall p \in P_{od}, \forall od, \forall s
\]

(9b)

\[
\frac{dx^p_{a_m}(s)}{ds} = e^p_{a_m}(s) - e^p_{a_{m-1}}(s), \forall a_m \in A_p, \forall p \in P_{od}, \forall od, \forall s
\]

(9c)

\[
\frac{dE^p_p(s)}{ds} = e_p(s), \forall p \in P_{od}, \forall od, \forall s
\]

(9d)
Mathematics in Transport Planning and Control

\begin{equation}
\sum_{\forall s} \sum_{\forall p \in P_{od}} E_p(s) = E_{od} \quad, \forall od \tag{9e}
\end{equation}

\begin{equation}
e_p(s) \geq 0 \quad, \forall p \in P_{od}, \forall od, \forall s \tag{9f}
\end{equation}

where $A_p$ is the set of all links on route $p$ and its cardinality is denoted by $|A_p| = M(p)$; $a_m$ represents the $m$-th link on the route where $m \in [1, M(p)]$. The notation $\tau^p_{a_m}(s)$ denotes the time of exit from link $a_m$ for traffic which enters route $p$ at its origin at time $s$; and $\tau^p_{a_m}(s)$ is the corresponding first derivative with respect to time.

Equations (9b) ensure the proper flow propagation along each route. Equations (9c) are the state equations that govern the evolution of link traffic. Equations (9d) define the relationship between inflow rate and the cumulative inflow $E_p(s)$ on each route and equations (9e) specify the total throughput $E_{od}$ between each origin-destination pair. Conditions (9f) ensure the positivity of the control variable $e_p(s)$. The travel time model satisfies FIFO structurally, hence we do not need to add any explicit constraint to this. This control theoretic problem involves state-dependent response. Its optimality conditions were first studied by Friesz et al. (2001) for inelastic equilibrium assignment. As an extension to Friesz et al., we derive the optimality conditions for elastic system optimal assignment as\(^1\):

\begin{equation}
e^*_p(s) \begin{cases} 
> 0 \Rightarrow \tilde{C}_p(s) + \lambda^p_{a_m}(s) - \gamma^p_{a_m}(s) = D^p_{od}[E^*_p(T)] \geq 0 \Rightarrow \tilde{C}_p(s) + \lambda^p_{a_m}(s) - \gamma^p_{a_m}(s) \geq D^p_{od}[E^*_p(T)] , \forall p \in P_{od}, \forall od, \forall s . \tag{10a}
\end{cases}
\end{equation}

It should be noted that the total travel cost $\tilde{C}_p(s)$ is different from the one $C_p(s)$ in equilibrium. The costate variable $\lambda^p_{a_m}(s)$ comes from the state equation of $x^p_{a_m}(s)$. For all links $a_m$ on route $p$, $\lambda^p_{a_m}(s)$ is governed by the following equation:

\begin{equation}
\frac{d \lambda^p_{a_m}(s)}{ds} = -\left(1 + f'[\tau^p(s)]\right)\frac{e^*_p(s)}{Q^p_{a_m}}, \forall a_m \in A_p, \forall p \in P_{od}, \forall od, \forall s . \tag{10b}
\end{equation}

In the optimal control problem, the costate variable $\lambda^p_{a_m}(s)$ represents the sensitivity of the optimal value of the objective function at and after the time $s$ with respect to a perturbation in the state variable $x^p_{a_m}(s)$ (Dorfman, 1969). The second multiplier $\gamma^p_{a_m}(s)$ comes from the flow

\(^1\) Readers can refer to Friesz et al. (2001) for the derivation of the optimality conditions of equilibrium assignment; or Chow (2005) for system optimal assignment.
propagation constraints \( (9b) \)\(^2\) associated with the outflow from each link. The costate \( \gamma_{a_{m}}^{p}(s) \) can be solved by the following set of recursive equations:

\[
\lambda_{a_{m}}^{p}[\tau_{a_{m}}^{p}(s)] - \gamma_{a_{m}}^{p}[\tau_{a_{m}}^{p}(s)] = \lambda_{a_{m}}^{p}[\tau_{a_{m-1}}^{p}(s)] - \gamma_{a_{m}}^{p}[\tau_{a_{m-1}}^{p}(s)] ; \quad (10c)
\]

\[
\lambda_{a_{m}}^{p}[\tau_{a_{m}(r)}^{p}(s)] = \gamma_{a_{M}(r)}^{p}[\tau_{a_{M}(r)}^{p}(s)] . \quad (10d)
\]

We can also give an economic interpretation for the costate variables. The costate variable \( \lambda_{a_{m}}^{p}(s) \) can be interpreted as the marginal cost of an additional traveller entering link \( a_{m} \) on route \( p \); while \( \gamma_{a_{m}}^{p}(s) \) is the marginal savings from a traveller leaving the link. Furthermore, subtracting the user equilibrium cost of a traveller from his/her marginal cost minus his/her marginal saving will yield his/her externality imposed to the system. This indeed is the optimal toll that the traveller has to pay, according to the marginal cost pricing principle.

5. EXAMPLE CALCULATIONS

We consider a network with two parallel routes connecting a single origin-destination pair. Route 1 has a free flow time 20 mins and a capacity 30 vehs/min; while the free flow travel time and the capacity of route 2 are 30 mins and 50 vehs/min respectively. The origin-specific cost is considered to be a monotone linear function of time with a slope -0.5. The destination cost function is piecewise linear which has no penalty for arrivals before the preferred arrival time 09:00, and increases with a rate 1.5 afterwards. The time incremental step \( \Delta s \) is set to be 1 min and the study horizon \( [0, T] \) is long enough such that all traffic can be cleared. An elastic demand function is added to specify the total throughput \( E_{od} \) generated, given the average travel cost \( C^{*} \) throughout the period. The demand function is defined as \( E_{od} = D \exp(\epsilon_0 C^{*}) \), where \( D \) and \( \epsilon_0 \) take the values of 6,190 (vehs) and -0.005 respectively.

5.1 Solution method for equilibrium assignment

Using (8), Mun (2002) proposed a solution method for the inelastic equilibrium assignment with a divided linear travel time model. We modify Mun’s algorithm for our travel time models and the elastic travel demand as follows:

**Step 0: Initialisation**

0.1. Select an initial equilibrium cost \( C^{*} \);

0.2. initialize the iteration counter \( n := 1 \);

\(^2\) It should be noted that the costate variable \( \gamma_{a_{m}}^{p}(s) \) will vanish if the outflow rate is fixed (e.g. when deterministic queuing model is adopted).
8 Mathematics in Transport Planning and Control

0.3. initialize \( e_{p}^{i}(k) := 0 \), for all routes \( p \) and all discretized time steps \( k \) from time 0 to \( K \), where \( K = T / \Delta t \) is total number of discretized time steps simulated;

0.4. initialise \( k := 0 \).

**Step 1: Network loading and equilibration**

1.1. Compute \( C_{p}(k) \) for all \( p \);

1.2. update the inflow profile as \( e_{p}^{n+1}(k) = e_{p}^{n}(k) - \pi d_{p}^{n} \) with an approximate direction \( d_{p}^{n} \) and step size \( \pi \). We adopt a second-order decent direction taken as \( d_{p}^{n} = \frac{\Omega_{p}^{n}}{\Omega_{p}^{n+1}} \) where

\[
\Omega_{p}^{n} = h'(k) - 1 + \left( 1 + f' \left[ \tau_{p}(k) \right] \right) \hat{\tau}_{p}(k)
\]

and

\[
\Omega_{p}^{n+1} = \frac{\partial \Omega_{p}^{n}}{\partial e_{p}^{n}(k)} = \left( 1 + f' \left[ \tau_{p}(k) \right] \right) \sum_{a} \delta_{p}^{n} \frac{1}{Q_{a}},
\]

where the indicator \( \delta_{p}^{n} \) equal to one if link \( a \) lies on route \( p \) and zero otherwise. The step size \( \pi \) is determined by a linear interpolation.

**Step 2: Stopping criteria**

2.1. If the convergence measure becomes sufficiently small, go to step 2.2; otherwise set \( n := n+1 \) and go to step 1.2;

2.2. if \( k = K \); go to step 2.3. Otherwise \( k := k + 1 \) and go to step 1.1;

2.3. check if \( E_{od} = \sum_{v} \sum_{t} e_{p}(k) = D \exp(\epsilon_{0} C^{*}) \). If yes, STOP; otherwise go to step 0.2 with

the updated equilibrium cost \( C^{*} := C^{*} - \left[ \frac{D \exp(\epsilon_{0} C^{*}) - E_{od}}{\epsilon_{0} D(C^{*}) - \frac{dE_{od}}{dC^{*}}} \right] \),

where \( \epsilon_{0} D(C^{*}) = \frac{dD(C^{*})}{dC^{*}} \) and \( \frac{dE_{od}}{dC^{*}} \) is the sensitivity of total throughput with respect to the equilibrium cost which can be calculated as in Heydecker (2002).

**5.2 Solution method for system optimal assignment**

The system optimal assignment can be solved as follows:

**Step 0: Initialisation**

0.1. Initialise a cost \( \tilde{C}^{*} \) at equilibrium;

0.2. initialise costates \( \lambda_{a}^{\lambda}(k) := 0 \) and \( \gamma_{a}^{\gamma}(k) := 0 \) for all routes \( p \) and all \( k \);

0.3. initialise \( e_{p}^{a1}(k) := 0 \) for all routes \( p \) and all \( k \);
0.4 set iteration counter $n := 1$;
0.5 initialise $k := 0$.

**Step 1: Network loading, optimization and equilibration**

1.1. Compute $\tilde{C}_p(k) + \lambda^a_p(k) - \gamma^a_p(k)$ for all $p$;
1.2. compute costate variable $\lambda^a_p(k)$ by solving equations (8b);
1.3. compute costate variable $\gamma^a_p(k)$ by solving (8c) – (8d) from $M(p)$ to 1 for all $p$;
1.4. update the control variable $e^{n+1}_p(k) = e^n_p(k) - \pi^* d^n_p$ with an approximate step size $\pi^*$.

We adopt a second-order decent direction taken as $d^n_p = \Theta^n_p/\Theta^n_p$, where

$$\Theta^n_p = h'(k) - 1 + f'[\tau^a_p(k)]\tau^a_p(k) + \dot{\lambda}(k) - \gamma(k)\tau^a_p(k)$$

and $\Theta^n_p = \frac{\partial \Theta^n_p}{\partial e^a_p(k)} = (1 + f'[\tau^a_p(k)]\sum_a \delta^a_p \frac{1}{Q_a}$.

The step size $\pi^*$ is determined by a linear interpolation.

**Step 2: Stopping criterion**

2.1 If the convergence measure becomes sufficiently small, go to step 2.2; otherwise set $n := n+1$ and go to step 1.2;
2.2. If $k = K$; go to step 2.3. Otherwise $k := k + 1$ and go to step 1.1;
2.3. check if $E_{od} = \sum_{v_p} \sum_{\forall a} e^*_p(k) = \tilde{D} \exp(e_0 \tilde{C}^*)$. If yes, STOP; otherwise go to step 0.2 with

the updated optimal cost $\tilde{C}^* = \tilde{C}^* - \left[ \frac{\tilde{D} \exp(e_0 \tilde{C}^*) - E^*_{od}}{e_0 \tilde{D}(\tilde{C}^*) - \frac{dE^*_{od}}{d\tilde{C}^*}} \right]$.

**5.3 Results**

Figure 3 shows the equilibrium assignments with deterministic queue and the whole-link models. For deterministic queuing model, the total traffic assigned to route 1 during times 07:52 and 08:57 is 2,358.5 vehs; while that to route 2 during times 07:52 and 08:38 is 2,112.5 vehs. With the same link travel times, link capacities and demand function, we also calculate the corresponding equilibrium flows for whole-link model. The traffic assigned to route 1 during times 06:50 and 09:00 is 2,167.26 vehs; while that to route 2 during times 07:10 and 08:03; and then from time 08:37 to 09:00 is 1,878.15 vehs. The traffic volume estimated by whole-link model is lower than that estimated by deterministic queue. This is due to the fact that the whole-link model will estimate a higher travel time and hence a higher total travel cost. In addition, the inflow profile estimated by the whole-link model is also more spread.
Figure 3 User equilibrium assignments

Contrast with deterministic queue, the period of assignment in system optimal is different from that in equilibrium. With the same traffic volume as in equilibrium, the period of assignment to route 1 shifts from [06:50, 09:00] to [07:11, 09:00]. For route 2, the first period of assignment shifts from [07:10, 08:03] to [07:41, 08:13]. The second period of assignment remains the same [08:37, 09:00] as before. It can be observed that the system optimal assignment, on the one hand, encourages late departures. On the other hand, it also has to maintain a certain amount of early departures to induce a high service rate for the departures at later times. After optimization, the estimated travellers’ surplus is increased by 508.11 veh-hr, from 13,484.70 veh-hr in user equilibrium to 13,992.81 veh-hr in system optimal.
We also plot the link traffic estimated by the whole-link traffic model in equilibrium and system optimal in Figure 5. Interestingly, yet importantly, the results show that the optimal assignment has to allow queuing. This implies that the analysis based on deterministic queue does not apply in general. Finally, to decentralise the system optimal flow, we need to impose the optimal tolls to the system which are shown in Figure 6. The optimal tolls for the two travel time models are substantially different. In particular, “negative toll” appears to encourage late departures for the whole-link traffic model. However, we will further investigate the causes and implications of this negative toll.

6. CONCLUDING REMARKS

This paper proposed a general framework for managing dynamic network traffic with plausible travel time models. The deterministic queuing model has been predominantly used in the literature for analysing dynamic road pricing. The analysis herein is more general and is applied to calculate the optimal flow and toll for Friesz’s whole-link travel time model. The significant differences that are established here between the two travel time models show that the analysis based on deterministic queue does not apply in general. This study provides the flexibility for choosing an appropriate traffic model and cost function. It also gives us a deeper understanding of the nature of optimal time-varying network flows and tolls. Future
work will include analyzing different tolling regimes and extending the present analysis to multi-destination networks with overlapping routes.

ACKNOWLEDGEMENTS

The continuing guidance and stimulating discussions from Professor Benjamin Heydecker and Dr. JD Addison are gratefully appreciated.

REFERENCES


