Hybrid Optimization: Control of Traffic Networks in Equilibrium

by

H.-N. Tan and
S.B. Gershwin

Massachusetts Institute of Technology
Laboratory for Information and Decision Systems
Cambridge, Massachusetts 02139

ABSTRACT

This paper considers the static traffic signal control problem in a network with the relaxation of the usual fixed flow assumption implicit in most such studies. Link flows in a traffic network are realistically assumed to be variable because given a set of network control parameters such as signal settings, drivers are free to choose among alternate paths. This is called "hybrid optimization" because it combines the usual notions of user equilibrium and system optimization. Necessary conditions for the optimum solution are derived and discussed. These conditions extend user equilibrium: not only is travel time equalized over utilized paths, but also other quantities related to the system-wide objective.

1. Introduction

The traffic control problem is related to the operational aspects of an automotive transportation system. The objective is to regulate traffic flows by using available control devices so that the existing facilities can be most efficiently utilized. A large amount of research has been undertaken. For example, dynamic control problems have been formulated for urban network traffic [1], [6], [7] and freeway corridor traffic [2], [3].

In this paper we focus on a special case where a steady state model of traffic flow is assumed. The vehicle traffic in a network is never at rest, but there are situations where certain quantities such as the rate of traffic demand and the traffic flow distribution can be assumed approximately constant for a relatively long period of time [4]. This kind of situation typically arises in the morning and the evening rush hours [11]. The steady state traffic model has traditionally been used to simplify the analysis of transportation networks [4], [5].

Fixed-time signal control policies have been widely used for traffic control due to their simplicity of implementation [6], [8], [10]. These are open-loop control policies where for a certain period of the day, the signals operate on a fixed cycle time, with fixed phases and offsets predetermined by some offline computation [10]. A large number of fixed-time signal optimization methods have been developed [8]. Typically in all these methods, optimal traffic controls are chosen to minimize total travel time assuming certain models for delay and traffic.

This research has been supported by the U.S. Department of Transportation under contract DOT-TSC-1456.

H.-N. Tan
S.B. Gershwin
behavior at a signalized intersection as a function of the control parameters.

Another very important assumption, which is seldom explicitly stated, is the assumption of fixed route choice of the drivers. This implies constant traffic volume on every link in the network under consideration. Under this assumption, the traffic on any particular link is constant regardless of the level of service offered by that link. This assumption is invalid in view of the fact that no individual driver can be prevented from taking an alternate route which could have been made more desirable, i.e. faster, by the implementation of a new policy. It seems intuitively convincing that in a fixed-time signal control system, drivers can learn to adapt their routes and speeds to advantage. In fact these redistributive effects of traffic resulting from implementation of an area traffic control policy have been confirmed in a series of field experiments conducted in Glasgow [12].

It is observed [12], [15] that the new traffic pattern indirectly induced by some "optimal" traffic control policy destroys the original optimality. It would thus seem desirable to periodically reoptimize the controls based on new survey information on the traffic distribution [10]. However, this process of updating controls has seldom been carried out more than once or twice in practice due to the amount of effort and resources involved [10]. On the other hand, it has also been shown that different signal timings induce different traffic patterns [13]. In a rather different approach, Allsop [14], recognizing the interdependence between signal timing plans and flow patterns, suggested the idea of using control schemes to influence drivers' route choice.

Given the fact that the system has little control over the route selection decisions of individual drivers, can one hope to achieve a flow distribution which is optimal from the system's point of view using the available control? Does the iterative reoptimization procedure necessarily lead to an optimal solution? Or more generally, given a certain predictive model of driver's route selection behavior, how should one go about choosing a set of controls which, together with the induced traffic pattern, is optimal with respect to a certain system cost criterion? This set of questions is of fundamental importance in transportation network planning.

The hybrid optimization problem has the following essential features. The objectives of the traffic authority and the drivers are different. On the system level, the problem for the traffic authority is to minimize some overall cost in the network, e.g. total travel time, or total fuel consumption. On the other hand, the individual driver wishes to minimize his trip cost in travelling through the network. Another important aspect in the hybrid optimization problem is the role of the individual drivers as independent decision makers in choosing among different available paths. This means that it is beyond the power of the traffic authority to establish link flow at any desirable volume. Consequently the capability of the traffic authority is limited to the command of traffic control devices only. In most cases, the capability of the traffic authority is further restrained because practical limitations dictate that the traffic authority can only excercise control over a subset of the network.

This paper represents an initial effort in this area of research. It should be emphasized that it does not lead immediately to a new design tool applicable for the solution of practical problems. The main
purpose here is to emphasize the importance of drivers' role as independent decision makers in transportation planning. We hope to provide a unified approach, a better understanding (at least qualitatively), an appropriate formulation and possible directions for algorithm development for this class of problem. A different formulation for a similar problem was proposed in [28], [29].

In section 2 we present a mathematical formulation of the hybrid problem. A heuristic procedure which has been proposed for the solution of the hybrid optimization problem is discussed in section 3. Necessary conditions of optimality are discussed in section 4, in particular the Extended Equilibrium principle which has an intuitively interesting interpretation. Two small networks are investigated numerically in section 5.

2. Formulation of the Hybrid Optimization Problem

An important and perhaps the most complicating issue in the hybrid optimization problem is the role of individual drivers as independent decision makers in choosing among different paths. In this section we first briefly review the flow distribution model to be used in this paper: user equilibrium. A mathematical formulation of the hybrid optimization problem is then presented.

2.1 Wardrop's First Principle: Equilibrium flow distribution

We assume in this paper that traffic distributes itself according to Wardrop's first principle [25] which states that "The journey times on all routes actually used are equal and less than those which would be experienced by a single vehicle on any unused route."

Suppose there are \( K \) origin-destination pairs and \( L \) links in the network under consideration. For the \( \ell \)th O-D (origin-destination) pair, let \( H^\ell \) be the total amount of traffic demand to go from the origin to the destination node. We denote the set of all paths connecting the \( \ell \)th O-D pair by \( P^\ell \). In this paper we use the notation \( |S| \) for the total number of elements in the set \( S \). Let \( T^\ell, h^\ell \) be the \( \ell \)th O-D pair path time and path flow vector respectively, each of dimension \( |P^\ell| \). Also let \( f \) and \( t \) be the link flow and link travel time vector respectively, each of dimension \( L \). The \( i \)th elements of \( h^\ell, h^j, T^\ell, T^j \), and \( f, t \), are respectively the path flow and trip travel time on the \( i \)th path of the \( \ell \)th O-D pair. The link flow and link travel time on link \( i \) are respectively denoted by \( f_i, t_i \), the \( i \)th element of \( f \), and \( t_i \), the \( i \)th element of \( t \). The path flow vectors for all O-D pairs are ordered to form a path flow vector, \( h \), which is of dimension \( \sum |P^\ell| \). Note that the link flow vector, \( f \), and the set of path flow vectors for all O-D pairs, \( \{ h^\ell \} \) are related by the equation

\[
f = \sum_{\ell=1}^{K} A^\ell h^\ell
\]

where \( A^\ell \) is the arc-path incidence matrix with dimension \( L \) by \( |P^\ell| \) for the \( \ell \)th O-D pair. The \((i,j)\) element of \( A^\ell \), \( a^\ell_{ij} \), is defined as follows:

\[
a^\ell_{ij} = \begin{cases} 
1 & \text{if link } i \text{ lies on the } j \text{th path of the } \ell \text{th O-D pair}; \\
0 & \text{otherwise}.
\end{cases}
\]
Suppose \( w \) is a vector of control parameters, e.g. green splits, ramp metering rates, etc. \( W \) denotes the set of feasible controls which defines the physical constraints on \( w \). The travel time on link \( i \), \( t_i \), is assumed to be a function of link flows, \( f_i \), and control parameters, \( w \), i.e., \( t_i = t_i(f, w) \). It is said to be separable if and only if it is independent of flows on other links, i.e.

\[
t_i(f, w) = t_i(f_i, w).
\]

The path and link travel time are related by the following equation

\[
T^\ell_{j}(f, w) = \sum_{i=1}^{L} a_{ij} t_i(f, w)
\]

or

\[
T^\ell_{j}(f, w) = (A^\ell)^T t(f, w)
\]

Wardrop's first principle can be expressed mathematically as follows:

\[
l_{k}\ell \sum \text{link-path flow relation} \quad f = \sum_{\ell=1}^{k} A^\ell h^\ell
\]

\[
h_i^\ell \geq 0 \quad \forall \ell \in \mathcal{P}, \quad \ell = 1, \ldots, K
\]

\[
\sum_{i=1}^{n} h_i^\ell = H^\ell \quad \ell = 1, \ldots, K
\]

Wardrop's first principle

\[
\begin{cases}
h_i^\ell > 0 \Rightarrow T^\ell_i(f, w) = \min_{j \in \mathcal{P}} T^\ell_j(f, w), \\
\ell = 1, \ldots, K
\end{cases}
\]

Definition A path \( i \) such that \( h_i^\ell > 0 \) for some \( i \) is a used path.

Definition A set of path flow vectors \( \{h^\ell\} \) is a feasible path flow if and only if \( \{h^\ell\} \) satisfies (4) and (5). The corresponding \( f \) which satisfies (3) is called a feasible link flow.

Definition A feasible path (link) flow is a user optimized or equilibrium path (link) flow if and only if (6) and (7) are satisfied.

For more detailed discussion on equilibrium flow, interested readers are referred to [5], [11], [16], [17], [18]. The problem of computing equilibrium flows has been shown [17], [18] to be equivalent to a convex programming problem under the assumption of separability and some additional mild assumptions. A recent result [19] using nonlinear complementarity theory has shown that an equilibrium flow pattern exists under very mild conditions:
hybrid optimization

2.2 Hybrid Optimization Formulation

The objective of the traffic authority is to minimize a system-wide cost function. For example, total travel time or total fuel consumption may be an appropriate cost criterion. In the case of total travel time minimization, the cost can be written as

$$J = \sum_{i=1}^{L} f_i t_i(f, w)$$

(8)

If the concern is total fuel consumption,

$$J = \sum_{i=1}^{L} f_i e_i(f, w)$$

(9)

where $e_i(f, w)$ is the amount of fuel consumed per vehicle in travelling through link $i$. In both cases, the system cost is a function of control parameters and link flows which can also be expressed as a function of path flows and control parameters using the link-path flows relation in equation (1), i.e. $J = J(f, w) = J(h, w)$.

It should be emphasized that the degree of freedom for the traffic authority is limited to the choice of control parameters from $W$. More importantly, $h$ is not an arbitrary feasible flow to be assigned by the traffic authority. Instead $h$ is required to satisfy a restrictive set of conditions which describes driver behavior. In this paper, Wardrop's principle is used for this purpose. Future research will be devoted to the use of other such descriptions, such as in [21].

Problem Statement

Minimize $J(h, w)$

subject to $h$ an equilibrium flow, $w \in W$ (10)

A mathematical optimization problem would be formulated in a straightforward manner from this problem statement if given any $w \in W$, $h$ were some known explicit function of $w$ or if $h$ were required to satisfy a set of equalities and inequalities. The problem is that neither of the two cases is true because for $h$ to be an equilibrium flow, $h$ is required to satisfy (3) - (7). A close examination of (3) to (7) shows that the dependence of $h$ on $w$ is not explicit. Furthermore the mathematical relations (6) and (7) are neither equalities nor inequalities. They are in fact two logical relations and hence they are not in a form to be posed as constraints to the optimization problem.

It can be shown that (3) to (7) can be transformed to the following equivalent form.

$$h_i^p \geq 0 \quad i \in P^p, \quad p = 1 \ldots K$$

(11)

$$\sum_{i=1}^{h_i^p = H_i^p} \quad p = 1, \ldots K$$

(12)

$$T_i^p(h, w) \geq \left( \sum_{j \in P_i^p} h_j^p T_j^p(h, w) / H_j^p \right)_{i \in P_i^p, \quad p = 1, \ldots K}$$

(13)
The hybrid optimization problem can now be stated precisely as follows:

\[
\begin{align*}
\text{minimize } & J(h,w) \\
\text{subject to } & (11)-(13).
\end{align*}
\]

It should be pointed out that the user optimization model for flow distribution is only one among the many available models [20], [21]. The user optimization model is used in this formulation for several reasons. It is a very common model and has been widely used in transportation planning [22], [23], [24]. Moreover all behavioral models are approximate ones and equilibrium flow has been shown [22] to be a reasonably good approximate of actual traffic distribution. However it should be emphasized that formulation (10) is not restricted to any particular flow distribution model.

3. A Heuristic Procedure

In this section we study a heuristic procedure which has been proposed and used in a number of studies [4], [13], [14], [12], [15]. This is an iterative procedure consisting of successive alternations between a signal optimizing program and an assignment program as shown in Figure 1. The assignment program computes an equilibrium flow assuming the

![Flowchart](image-url)

**Figure 1:** The Heuristic Procedure
control parameters are fixed. The signal optimizing program computes a set of optimal signal settings with respect to some system cost assuming flows to be fixed. The procedure is initiated by a guess of the optimal control parameters and proceeds by iterating between the two programs until a stopping criterion is satisfied.

As a computation method, the heuristic procedure is attractive, especially for large-scale network applications. Both the components of this procedure, the signal optimizing and assignment problems, have been studied extensively and well developed computer software is available. However, various aspects of this procedure have not been closely examined. For example, does the procedure converge, what are the properties of the result?

In this section we establish the fact that the heuristic procedure does not necessarily converge to the optimal hybrid solution by using a simple counterexample. Consider the network in Figure 2 with link

\[ t_1 = 15 + 2f_1, \quad t_3 = 15 + 2f_3, \]
\[ t_4 = 50 + f_4, \quad t_5 = 50 + f_5. \]

Figure 2: Network for Example 1
Link 2 is under the control of the traffic authority and for simplicity we assumed that the control is in the form of delay imposed on the traffic passing through link 2. Therefore
\[ t_2 = 10 + f_2 + w, \]
where \(w\) is the imposed delay. Ten units of traffic flow are required to go from node 1 to node 2.

**Assignment Program:** Let \(x\) be the path flow along path \((1, 3, 4, 2)\). Because of symmetry the path flows along \((1, 3, 2)\) and \((1, 4, 2)\) are the same and equal to \((10 - x)/2\). Figure 3 shows \(x\) at equilibrium as a function of \(w\).

![Figure 3: Flow along (1,3,4,2)](image)

**Signal Optimizing Program:** Assuming that the traffic authority wishes to minimize total travel time. The problem of the control optimizing phase of the heuristic procedure can be shown to be the following.
\[
\min_{w} \quad xw + 2.5((x-3)^2 + 311)|_{x \text{ fixed}}
\]
subject to \(w \geq 0\)
The solution of this problem can be shown to be
\[
w = 0 \quad \text{for} \quad x > 0
\]
\[
w > 0 \quad \text{for} \quad x = 0
\]
(14)

We apply the heuristic procedure for this example with initial guess of \(w = 10\). The result of the heuristic procedure is listed in Table 1 which is constructed using (14) and Figure 3. The converged

<table>
<thead>
<tr>
<th>STEP</th>
<th>SYSTEM (w)</th>
<th>DRIVERS (x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>INITIALIZATION</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>ASSIGNMENT</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>CONTROL OPTIM.</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>ASSIGNMENT</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>CONTROL OPTIM.</td>
<td>0</td>
<td>8</td>
</tr>
</tbody>
</table>

CONVERGED SOLUTION: \(w = 0, x = 8\)

Table 1: Result of the Heuristic Procedure
solution is \( w = 0, x = 8 \).

Figure 4 shows the system cost as a function of \( w \). It is clear that the lowest cost is achieved with \( w > 20 \) and Fig. 3 implies that the optimal flow on link 2 is \( x = 0 \). That is, it is best not to use this link, and there should be a sufficiently large control delay imposed so that no driver chooses to travel on it. This is thus an example of Braess' paradox [5].

The heuristic procedure, however, has done the opposite and converged to the worst possible solution. This is because the control \( w \) is chosen as a function of the existing assignment, and not in anticipation of the next assignment.

4. Extended Equilibrium Principle for Hybrid Optimization

We present in this section an extended equilibrium principle for the hybrid optimization problem without proof. This is an interpretation of some of the Kuhn-Tucker necessary conditions of optimality. For detailed information on the derivation, the readers are referred to [26].

Suppose \( \{f^*, w^*\} \) is the optimal hybrid solution. Then there exists some \( \{\lambda^L \in \mathbb{R}, L = 1, \ldots, K\} \) and \( \{\Pi^L \in \mathbb{R}^+, L = 1, \ldots, K\} \) which are related to the Lagrange Multipliers of the hybrid optimization problem, such that the following statements are true.

1. The trip times along all used paths between the \( L \)th O-D pair are the same and equal to \( T^L_{\text{min}} \). Any path having trip time greater than \( T^L_{\text{min}} \) carries no flow.
Let 
\[ \psi_i^j = \frac{\partial}{\partial f_i} + \lambda^j \tau_i^j + \sum_{j=1}^{K} (\lambda^j f^j - \Pi^j)^T \frac{\partial}{\partial f_i} \]
for each link and each O-D pair \( \ell \), evaluated at \((f,w) = (f^*,w^*)\). Let \( M_{\ell}^j = \sum_{j=1}^{K} a_{ij}^j \psi_i^j \) be a pseudo-cost along the \( j \)-th path of the \( \ell \)-th O-D pair. Note that \( f^i = A^i_{hi} \) is the link flow vector for the \( i \)-th O-D pair.

We have the following additional equilibrium principle:

(2) The pseudo-costs along all paths between the \( \ell \)-th O-D pair are the same and equal to \( M^\ell_{\min} \). Any path having a pseudo-cost greater than \( M^\ell_{\min} \) carries no flow.

The first statement is Wardrop's first principle. The second has been written in a way to emphasize its similarity with that principle. The quantity \( M^\ell_{\min} \) can be thought of as the marginal cost of the \( j \)-th path for the \( \ell \)-th O-D pair. While statements (1) and (2) seem to provide opportunities for research into optimization techniques, it must be remembered that \( M^\ell_{\min} \) involves the vectors \( \lambda^j \) and \( \Pi^j \), which are unknown.

5. Numerical Examples

It has been shown in section 3 that the intuitively appealing action-reaction heuristic procedure can converge to the worst possible solution. The main difficulty with the heuristic procedure is that in the control optimizing phase, controls are chosen without taking the reactions of the drivers into consideration. In this section we describe two simple examples to demonstrate that the formulation of the hybrid optimization is indeed well defined. Details appear in [26].

5.1 Example 1

We solve the same problem as in section 3 using a general nonlinear programming technique, with the same initial guess of \( w = 10, x = 4 \). A solution of \( w = 20 \) and \( x = 0 \) is obtained.

We list in Table 2 the flow, travel time, and pseudo-cost along each path to illustrate the Extended Equilibrium principle.

<table>
<thead>
<tr>
<th>path</th>
<th>flow</th>
<th>travel time</th>
<th>pseudo-cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 : (1,4,2)</td>
<td>5</td>
<td>80</td>
<td>95</td>
</tr>
<tr>
<td>2 : (1,3,2)</td>
<td>5</td>
<td>80</td>
<td>95</td>
</tr>
<tr>
<td>3 : (1,3,4,2)</td>
<td>0</td>
<td>80</td>
<td>100</td>
</tr>
</tbody>
</table>

Paths 1 and 2, each carrying a positive amount of flow (5 units), have equal trip time, \( T^* = 80 \). Path 3, which has no flow, has a travel time not less than \( T^* \). The second part of the Extended Equilibrium Principle, i.e., the principle of equalization of pseudo-costs, is also satisfied in Table 2. Paths 1 and 2, each carrying five units of flow, have a same pseudo-cost of \( M^\ell_{\min} = 95 \). Path 3, which has a pseudo-cost of \( 100 > M^\ell_{\min} = 95 \), carries no flow.

5.2 Example 2

Consider the network as shown in Figure 5. Node 3 is a signalized intersection. We denote the green split facing link \((1,3)\) by \( g \). The travel time on all links is modelled as a fourth power polynomial of the link flow [27]

\[ t_i(f_i) = t_o (1.0 + 0.15(\frac{f_i}{\text{capacity}})^4) \]
The capacity is taken to be 1500 vehicles/hour per lane. All links except link (4,5) are assumed to be single-lane. We use the Webster formula [15] with fixed cycle time of 1.0 minute for the waiting time at the signalized intersection i.e., at links (1,3) and (2,3). The traffic demands are 800 veh./hr. for each of the following O-D pairs: 1 to 5, 1 to 6, 2 to 5, 2 to 6.

This example is solved using the formulation presented in section 2 by a general nonlinear programming technique. The optimal solution obtained is $g^* = .215$ with the minimum cost (total travel time) equal to 880.72 vehicle-hour/hr. The optimality of $g^*$ is verified using the results of a series of user-optimized flow patterns computed at various values of $g$, the green split. The associated system cost is shown in Fig. 6, which demonstrates the optimality of $g^*$.

We list in Table 3 the flow, travel time, and pseudo-cost along each path to demonstrate the validity of the Extended Equilibrium Principle.

<table>
<thead>
<tr>
<th>O-D Pair</th>
<th>Path</th>
<th>Path Flow</th>
<th>Path Time (Min)</th>
<th>Path Pseudo-cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1</td>
<td>1(1,4,5)</td>
<td>800.00</td>
<td>13.61</td>
<td>2.55204x10^{-1}</td>
</tr>
<tr>
<td>1 2</td>
<td>2(1,3,4,5)</td>
<td>0.0</td>
<td>15.70</td>
<td>2.73347x10^{-1}</td>
</tr>
<tr>
<td>2 1</td>
<td>1(1,4,5,6)</td>
<td>505.72</td>
<td>14.81</td>
<td>4.39999x10^{-1}</td>
</tr>
<tr>
<td>2 2</td>
<td>2(1,3,6)</td>
<td>294.28</td>
<td>14.81</td>
<td>4.40000x10^{-1}</td>
</tr>
<tr>
<td>2 3</td>
<td>3(1,3,4,5,6)</td>
<td>0.0</td>
<td>16.90</td>
<td>4.81376x10^{-1}</td>
</tr>
<tr>
<td>3 1</td>
<td>1(2,3,4,5)</td>
<td>222.30</td>
<td>19.26</td>
<td>2.18417x10^{2}</td>
</tr>
<tr>
<td>3 2</td>
<td>2(2,5)</td>
<td>577.70</td>
<td>19.26</td>
<td>2.18417x10^{2}</td>
</tr>
<tr>
<td>3 4</td>
<td>1(2,3,4,5,6)</td>
<td>0</td>
<td>20.47</td>
<td>3.53031x10^{-1}</td>
</tr>
<tr>
<td>4 4</td>
<td>2(2,5,6)</td>
<td>0</td>
<td>20.47</td>
<td>3.53032x10^{-1}</td>
</tr>
<tr>
<td>4 4</td>
<td>3(2,3,6)</td>
<td>800.00</td>
<td>18.37</td>
<td>3.34892x10^{-1}</td>
</tr>
</tbody>
</table>

Table 3 Path Times and Pseudo-costs for Example 2

Note that certain paths are unutilized because both the travel times and pseudo-costs are too great. This differs from Example 1 in which a path is unutilized only because of the pseudo-cost.
5.3 Solution by the Heuristic Procedure

In the second example, we also applied the Heuristic Procedure to minimize travel time. An initial value of 0.6 is assumed for the green split $g$. Figure 8 shows the values of $g$ from iteration to iteration. The final converged solution is $g = 0.0$, which is not the optimal solution of the Hybrid Optimization Problem as shown in Fig. 6.

6. Conclusion

It is the main purpose of this paper to emphasize the importance of drivers' behavior in transportation planning. Numerical examples have been presented to show that proper evaluation of any control policy cannot be made without taking the reactions of the drivers into consideration. Numerical experience shows that the heuristic procedure leads to wrong solutions in both the examples. A practical implication is that if this procedure is implemented in real life as suggested in [10], [15], the amount of excess cost accrued over a long period of operation can be substantial.
The numerical examples presented also show that the formulation of the hybrid optimization problem is indeed well defined. We introduce an extended notion of the user equilibrium principle for the hybrid optimization problem and it is verified using two numerical examples.

References


7. P. Jensch, "A Dynamic Phase Optimization and a Decision Table Method for Control of Vehicle Actuated Traffic Signals", in


