Traffic flow patterns in AHS networks: system and user optimals 1

Gabriel Gomes2, Luis Alvarez3 and Roberto Horowitz4

1,3 Department of Mechanical Engineering, University of California, Berkeley, CA 94720-1740, USA
2 Instituto de Ingeniería Universidad Nacional Autónoma de México, 04510 Coyocán DF, México

Abstract

In this paper a method of traffic control for the link-network layers of an automated highway system (AHS) is presented. The goal of the method is to provide a time sequence of activities and on-ramp metering rates which guarantees traffic safety, while preserving a balance of user and system benefit. The approach is based on a static principle of vehicle conservation and on the notion of activities, which dictate the behavior of traffic flows. The problem is posed as a set of linear programming problems, which are solved sequentially.

1 Introduction and Background

In [6], the author proposes a decomposition of the control tasks of an automated highway system (AHS) into four layers; network, link, coordination and regulation, which are in order of decreasing abstraction. Under this architecture, each layer has a specific control objective, which influences and depends on the layers above and below it. The AHS is equipped with a single network layer controller. It views the system as a connected set of nodes and links, and is responsible for assigning appropriate routes for each vehicle entering the system. Each link layer controller focuses on a smaller portion of the highway. Its task is to ensure a smooth flow of vehicles by taking safety considerations into account, and by adapting to incidents. Also, the link layer controller may adjust on-ramp access rates, velocities and maneuvers to quickly respond to changes in local conditions, such as the onset of congestion or transit of emergency vehicles [2]. The lower layers of the hierarchy: coordination and regulation, are located onboard, and are dedicated to controlling individual vehicles and coordinating maneuvers among groups of vehicles in a safe and efficient manner.

In this paper we focus on the link-network layers. At this level of abstraction, individual vehicles are not considered, but rather types of vehicles, which are characterized by their origin/destination pair. Vehicle types can also distinguish other features, such as trucks versus cars, or leaders versus followers in a platoon. The main goal of these layers is to provide an activity plan and a velocity plan for each vehicle type utilizing the system, as well as an entry plan for each on-ramp. The activity plan consists of proportions of vehicles of a given type that should perform each activity or maneuver (e.g. lane change, split, join). Thus, a feasible activity plan should direct the flow of traffic of each vehicle type from origin to destination, while satisfying capacity and operational constraints. This framework for the higher layers of the AHS was first presented in [3] and subsequently considered in [1].

The link layer controller, as presented in [4], is divided into feedforward and feedback portions. The task of finding optimal plans is pertinent to the feedforward controller, while the feedback controller measures and stabilizes the highway state. In this work, we pose the general feedforward problem and seek numerical solutions of two types: 1) Steady state solutions 2) Finite horizon dynamic solutions when the velocity plans are prescribed. Platooning is not considered in order to reduce the set of activities to three: lane change left, lane change right, and straight. However, extending the approach to platooning behaviors only increases the dimension and not the structural complexity of the problem.

One of the difficulties in the design of traffic assignment plans is to achieve balance between system benefit and equity among users. Considering only global performance measures such as total flow or total travel time may cause a minority of users to be burdened with long waits, in order to alleviate flow for the majority. We address this issue by defining a measure of fairness of a solution as the variance in the set of queueing times experienced by drivers at every time. Thus, a solution is perfectly fair if, at every time, the expected queueing times are equal in all entries to the AHS. We investigate the tradeoff of fairness and throughput in the case of static solutions.
2 Notation

This paper follows the notation and modelling paradigm presented in [3] and [1]. Each AHS link is modelled as a set of interconnected sections, each representing a one lane, half mile stretch of highway (see Fig. 1). Traffic flow between sections occurs either longitudinally, or laterally via lane change. However, the assumption that sections are connected downstream and upstream to at most one other section. Thus, traffic flow bifurcations require prior lane change maneuvers in an upstream section.

$S$ will denote the set of all sections in the link. $R$ and $E$ are the subsets of on-ramp and exit sections. The defining parameters of each section are its length, $L(i)$, and its maximum capacity, $TS(i)$ measured in ($m \times s$). $V$ denotes the set of all vehicle types. In the numerical examples presented below, we only encode origin/destination information in the types, and denote the set of simulation times as $T$. We consider three possible activities or maneuvers: straight, lane change left, and lane change right, and denote the number of vehicles of type $j$, in section $i$ and at time $t$, performing each maneuver as: $N_s(i,j,t)$, $N_l(i,j,t)$, $N_r(i,j,t)$. We assume that a known flow of vehicles, $Q_a(i,t)$ (measured in veh/sec), enters queue $i \in R$. Of these, a proportion $\mu(i,j,t)$ is of type $j$. At each on-ramp there is a queue of $Q(l(i,t))$ vehicles, which are let onto the AHS at a rate of $Q(i,t)$ (veh/hr).

Within each section, all vehicles are assumed to travel at the same velocity $v(i,t)$, regardless of type.

3 Traffic Model

1) Vehicle conservation: The density variables, $N_s(i,j,t)$, $N_l(i,j,t)$ and $N_r(i,j,t)$, obey the following conservation principle:

$$N_s(i,j,t + \Delta t) + N_l(i,j,t + \Delta t) + N_r(i,j,t + \Delta t) = [N_s(i,j,t) + N_l(l(i,t),j,t) + N_r(r(i,t),j,t)](1 - \tau(i,t)) + [N_s(p(i),j,t) + N_l(l(p(i)),j,t) + N_r(r(p(i)),j,t)]$$

$$\tau(p(i),t) + \mu(i,j,t)Q(i,t)\Delta t,$$

where $\tau(i,t) := v(i,t)\Delta t/L(i)$ is the proportion of vehicles which move to downstream sections over $\Delta t$. $l(i)$, $r(i)$, and $p(i)$ are the left, right and previous (upstream) sections of $i$ (Fig. 1). In this equation, the left hand side is the total number of vehicles of type $j$, in section $i$, at time $t + \Delta t$. The flow of vehicles can be understood as occurring in two stages. First, all lane change maneuvers are completed, and become straight maneuvers. This stage hinges on the assumption that the lower hierarchical layers are able to complete all lane changes within a single section, and in less than $\Delta t$ time. For this to hold, it is necessary that there is sufficient available space in the adjacent lane. Thus, a restriction is placed on the amount of space required by feasible activity plans. This capacity constraint is explained further below. The second stage consists of vehicle flow into downstream sections. For vehicle conservation to hold, $v(i,t)$ and $\Delta t$ must be chosen such that $\tau(i,t) \in [0,1]$ $\forall i,t$.

2) Queue conservation: The dynamics of the queues follows a similar conservation equation:

$$Q_l(i,t + \Delta t) = Q_l(i,t) + (Q_a(i,t) - Q(i,t))\Delta t$$

where $Q(i,t)$ is the number of vehicles waiting in the queue, and $Q(i,t)$ the flow of vehicles being granted access to the AHS.

3) Capacity constraint: The capacity constraint guarantees that maneuvers are completed successfully before the vehicle exits the section by assigning to each maneuver a required space. It has been observed in micro-level simulations of the AHS that maneuvers such as lane change may be severely delayed by coordination layer protocols when there is insufficient space in the adjacent lane. In these cases, even though safety is guaranteed by the lower layers, inability to complete maneuvers on time may lead to missed exits and unwanted disturbances in link layer control. The concept of highway service used in [3] (measured in $m \times s$), accounts for both the time and the space needed to complete each maneuver gracefully, within a single section. The average space required for a particular maneuver may depend on vehicle type (e.g., when type distinguishes trucks from cars), and is an increasing function of the section velocity. Also, maneuvers such as lane change use space in more than one lane. Space requirements are denoted $\lambda ij [ms/veh]$, where $i$ refers to the maneuver ($s$, $l$ or $r$), and $j$ to the section ($c$ or $d$, depending on whether the space is occupied in the current or adjacent lane).

The capacity constraint for the non-platooning, fixed velocity case is

$$\sum_{j \in V} [N_s(i,j,t)\lambda xc(j) + N_r(i,j,t)\lambda rc(j) +$$

$$N_l(i,j,t)\lambda lc(j) + N_r(l(i),j,t)\lambda sa(j) +$$

$$N_l(r(i),j,t)\lambda ra(j)] \leq TS(i),$$
where $TS(i)$ is the total amount of space in section $i$ (e.g. $L(i)\Delta t$ under normal conditions).

4) OD constraint: A feasible activity plan, in addition to satisfying conservation and capacity constraints, should also direct vehicles to their desired exit sections. This condition is expressed by the following set of equality constraints:

$$N_s(i,j,t) = 0 \quad \forall t \quad i \in E \quad i \neq dest(j)$$  \hspace{1cm} (4)

where $dest(j)$ is the destination section for type $j$.

5) No lane skipping: Vehicles are not allowed to enter and exit a particular section by lane changes. This is to account for the capabilities of individual vehicles and the discrete time of the model, and implies the following constraint on the aggregate traffic plan:

$$N_e(i,j,t) + N_t(i,j,t) \leq \left[ N_s(p(i),j,t-\Delta t) + N_t(p(i),j,t-\Delta t) + N_s(pr(i),j,t-\Delta t)\right] t(p(i),t-\Delta t) + N_s(i,j,t-\Delta t)(1 - t(i,j,t-\Delta t)).$$  \hspace{1cm} (5)

6) Exit capacity: The flow out of the AHS will typically be limited by the capacity of the off-ramps and connecting manual lanes. This imposes the following condition on feasible plans:

$$\sum[N_s(i,j,t) + N_t(l(i),j,t) + N_t(r(i),j,t)] t(i) < EC(i).$$  \hspace{1cm} (6)

where $EC(i)$ is the maximum exit flow allowed in off-ramp $i$.

4 General problem formulation

We wish to compute activity, velocity and entry plans which achieve a high level of performance of the AHS. In our formulation, this means finding positive $N_s$, $N_t$, $N_r$, $v$, $Ql$, and $Q \quad i \in S$, $j \in V$, and $t \in T$, which minimize the total time spent in the system [5].

$$\min \left\{ \sum_i \sum_j \left[ N_s(i,j,t) + N_r(i,j,t) + N_t(i,j,t) \right] + \sum_i Ql(i,t) \right\}$$  \hspace{1cm} (7)

while satisfying constraints 1-6.

Two approaches are taken to handle this problem, both assuming that the velocity plan is given (e.g. $v(i,t) = v_{max}$, the maximum permitted velocity). Notice that the velocity is the only variable which enters nonlinearly into the equations (through the $\lambda$'s). Thus, this assumption converts the problem into a linear programming one, for which numerical solutions can be computed with relative ease. The simplified problem can be casted into a standard matrix form:

$$\min(w^T X) \quad \text{subject to: } AX \leq B; \quad X \geq 0$$  \hspace{1cm} (8)

where $X$ is the state vector containing the density, velocity, and entry flow variables, and $A$, $B$, and $w$ are appropriately sized matrices. The $\leq$ and $\geq$ signs express that each row is either an equality or an inequality constraint. Our first approach is to find static solutions, i.e. solutions in which none of the variables depend on time.

5 Static solutions

In searching for static solutions to the general problem, we are assuming that the inlet and capacity conditions of the AHS remain constant for a sufficiently long period of time. This assumption is justified when the demands on the highway are low and we do not expect formation of large queues. Nevertheless, we do not make the assumption of low inlet demands, and search for solutions which are fair, in the sense that they the queueing time evenly among all on-ramps.

To obtain static solutions, time dependencies are removed from all variables. Also, the $Ql$ state is removed and its conservation equation replaced by an upper bound on the metering rate: $Q(i) \leq Qa(i)$. The cost function (7) is no longer valid since it involves both time and queue lengths, and simply minimizing the number of vehicles on the highway leads to the trivial solution of zero input. Instead, the problem is solved in three stages: flow maximization, variance minimization, and solution improvement.

Step 1: Flow maximization. The first stage consists in finding the maximum achievable flow through the link. This is the solution to the following linear maximization problem:

$$\max(\sum_i Q(i)) \quad \text{subject to modified constraints} \quad 1,3,4,5,6$$

If the inlet demands $Qa(i)$ are low, the solution to this problem is simply $Q(i) = Qa(i) \forall i$. In this case, the solution is optimal and we are done. However, demands exceeding the link total capacity will lead to the formation of queues in one or more of the on-ramps. This may produce solutions which are unsatisfactory because they involve large differences in the queueing time among different entries. Thus, we have attained a system optimal, in the sense that the AHS throughput is maximal, and yet some users experience unfairly large travel times. In the second stage, we allow a limited sacrifice of system optimality to improve on user fairness.
Step 2: Variance minimization. The normalized inlet error is defined as:

\[ \tilde{Q}(i) = (Qa(i) - Q(i))/Qa(i) \]

For static solutions, the expected queueing time is proportional to \( \tilde{Q} \). We therefore redefine fairness in terms of this quantity and seek to minimize the variance in the set of normalized inlet errors:

\[
\min(\sum_{i \in \mathcal{I}} [\tilde{Q} - \tilde{Q}(i)]^2) = \min(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{V}} \omega_{ij} Q(i)Q(j))
\]

where \( \tilde{Q} = \frac{1}{P} \sum_{i \in \mathcal{I}} \tilde{Q}(i) \)

\[
\omega_{ij} = \begin{cases} 
P^{-1} & \text{if } i = j \\
\frac{PQa(i)}{Qa(i)Qa(j)} & \text{if } i \neq j 
\end{cases}
\]

and \( P \) is the number of on-ramps. We require the solution to be close to the system optimal by adding the constraint:

\[
\sum_{i \in \mathcal{O} \cup \mathcal{E}} Q(i) \geq \beta Q_{\text{max}}
\]

where \( Q_{\text{max}} \) is the value of the optimal cost from Step 1 and \( \beta \in [0,1] \) is a factor which determines the maximum percentage of flow to be sacrificed.

To illustrate this approach, consider the link of Figure 2, with two vehicle types, A and B, each with inlet demands of \( Qa = 1 \text{ veh/sec} \). Notice that there is only one admissible route for type B (due to the lane skipping constraint), while vehicles of type A have two possibilities: a long route (through sections 7,8,9) and a short route (through 5,6). Assuming that the link has a maximum flow capacity of 1 veh/sec, any feasible solution with \( Q(A) + Q(B) = 1 \) is a solution to Step 1. Step 2 gives preference to the zero variance solution: \( Q(A) = Q(B) \), while guaranteeing a minimum level of link usage determined by \( \beta \). This stage is guaranteed to have at least one solution, namely, the solution of Step 1.

Step 3: Improvement. The final step is essentially the same as Step 1: maximization of the total throughput, with the added condition that all queueing rates are bounded below by the solution of Step 2. In other words, we attempt to increase the inlet flow in one or more entries without decreasing any. In our example, with \( \beta = 0.8 \), \( Q(A) = 0.4 \) solves Step 2. In Step 3, one or both rates may be increased to 0.5 veh/sec.

The search space for Steps 2 and 3 can be reduced by computing a lower bound on the total flow through the link:

\[
\sum_i Q(i) \geq Q_{\text{min}}
\]

The value of \( Q_{\text{min}} \) is found by solving an intermediate linear programming problem in which achievable throughput is maximized, while requiring zero variance in the inlet errors:

\[
Q_{\text{min}} = \max(\sum_i Q(i))
\]

subject to:

\[
\tilde{Q}(i) = \tilde{Q}(j) \quad \forall \quad i \in \mathcal{I}, j \in \mathcal{V}
\]

Any solution yielding a total flow less than \( Q_{\text{min}} \) can be discarded since \( Q_{\text{min}} \) is achievable with zero variance. With \( Q_{\text{min}} \) we define \( \beta_{\text{min}} := Q_{\text{min}}/Q_{\text{max}} \). In Step 2, the parameter \( \beta \in [\beta_{\text{min}},1] \) can be adjusted to give preference to variance minimizing \( (\beta = \beta_{\text{min}}) \) or flow maximizing \( (\beta = 1) \) solutions.

6 Finite horizon dynamic solutions

An obvious drawback of the static approach is that it does not account for temporal variations in the demands on the AHS. However, it provides a simple technique for generating user/system balanced plans. In this section we illustrate the usefulness of dynamic solutions of the general problem. Notice that Eq. 7 is a global cost and therefore does not guarantee an equitable solution.

We assume that there exists a finite time \( t_f \) such that the given velocity plan \((v(i,t))\) and inlet demands \((Qa(i,t))\) become constant after \( t_f \). We also assume that the steady state values of the demands can be accomodated by the AHS without unbounded growth of any queues. This allows us to define a time \( t_f \) large enough for the queues to empty and the state of the link to become close to steady state. In order to obtain sensible solutions, the problem must be solved over a time window \([0,T]\), for \( T \geq t_f \).

7 Numerical examples

Following, we give examples of numerical solutions to the problems presented in this paper. All cases refer to the link shown in Fig. 2, with sections measuring 500 m, and a constant and uniform traffic velocity of 50
mph. Straight maneuvers are assumed to require 50 m per vehicle, while lane change maneuvers require 25 m in the current lane and 50 m in the adjacent lane.

The static problem was solved with inlet demands of $Q_a = 1 \text{veh/h}$ in both inlets. Step 1 provided a maximum steady state throughput for the link of 0.62 veh/sec. Thus, the link cannot support the demand of 2 veh/s without forming queues. The intermediate step gave a maximum zero-variance throughput of 0.56 veh/sec which implies $\beta_{min} = 0.56/0.62 = 0.89$. Fig. 3 shows solutions of steps 2 and 3, for values of $\beta$ ranging between $\beta_{min}$ and 1. The tradeoff between user and system optimal static solutions is apparent.

Figure 4 illustrates results of the dynamic problem. The top three graphs show the queue demands, lengths and access rates, related to types A and B, which enter through sections 0 and 1 respectively. The remaining graphs show the evolution of the total number of vehicles of each type occupying key sections in the AHS. Notice that vehicles of type A begin by taking the shorter route, but switch to the longer after approximately 150 sec., when vehicles of type B saturate the juncture (section 5). The use of the longer routes is evident in the increase in density in sections 2 and 7. Queue formation begins around 300 sec. and both queues reach a maximum length of approximately 40 vehicles, and then empty out. Notice that, even though both queues have comparable sizes, the queueing time may vary largely between queues due to different access rates.

8 Conclusions

In this paper we found that the aggregate quantities used to describe the AHS at the link/networks level are constrained by a set of equations which are linear in the vehicle densities and on-ramp access rates, and nonlinear in the traffic velocity. Assuming constant velocity and steady state, we found solutions which account for both system and user costs by solving a sequence of linear and quadratic programming problems. An immediate extension to this work is to incorporate a measure of fairness into the dynamic solution. Thus, we may sacrifice system optimality in order to increase inlet queue rates in trouble on-ramps. Other extensions include consideration of platooning maneuvers, such as join and split, and solving for section velocities. The former implies an increase in the dimension of the optimization problem, and addition of constraints specifying transitions between maneuvers, while the latter is a nonlinear programming problem.

Acknowledgement: The authors wish to thank Pablo Durango and Maksim Oks from the IEOR Dept. at U.C. Berkeley for their help with the numerical results.

References