

Program on Advanced Technology for the Highway
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA AT BERKELEY

Sketch of an IVHS Systems Architecture

Pravin Varaiya
Steven E. Shladover

PATH Research Report
UCB-ITS-PRR-91-3

This work was performed as part of the Program on Advanced Technology for the Highway (PATH) of the University of California, in cooperation with the State of California, Business and Transportation Agency, Department of Transportation, and the United States Department of Transportation, Federal Highway Administration.

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification, or regulation.

October 17, 1990
revised February 2, 1991

ISSN 1055-1425

Sketch of an IVHS Systems Architecture

Pravin Varaiya and Steven E. Shladover
PATH/ITS
University of California, Berkeley CA 94720

17 October 1990
revised 2 February 1991

¹Work supported by PATH (Program on Advanced Technology for the Highway), Institute of Transportation Studies, University of California, Berkeley, and NSF Grant ECS-8719779.

Contents

1	Introduction	2
2	Organization of traffic	3
3	Control tasks	7
3.1	Route and flow control	8
3.2	Path and congestion control	9
3.3	Platoon maneuvers	10
3.4	Feedback control	11
4	IVHS control architecture	12
5	Reference models and open architecture	14
6	Communication requirements	18
7	IVHS architecture for partial automation	19
8	Full or partial automation	22
9	Conclusions	24

List of Figures

1	A route is a sequence of segments	4
2	A path within a segment	5
3	Definition of platoon	6
4	IVHS control architecture	13
5	Link layer reference model	15
6	Platoon layer reference model	15
7	Regulation layer reference model	16
8	Physical layer reference model	17

9	Partially automated IVHS control architecture	21
---	---	----

Abstract

The use of control and communication technologies in vehicles and in the highway in the form of an Intelligent Vehicle/Highway System or IVHS is an approach that promises to improve capacity without building new roads. The paper presents a sketch of a system architecture for IVHS to carry out its control and management tasks. The discussion is divided into two parts. In the first part we focus on the tasks of a 'fully automated' IVHS in which the driver has little control. The tasks are differentiated along four dimensions:

- function – the functions range from stabilizing individual vehicles along nominal trajectories to adapting traffic flows to changing demands;
- time scale – the frequency of decisions and responses varies from under 1 s for continuous control of vehicles to several hours for network flow optimization;
- spatial scope – the impact of a control action can vary from a single vehicle to the traffic in the entire network;
- information span – satisfactory accomplishment of the task will require information ranging from that referring to a single vehicle to that which spans system-wide flows.

We outline an architecture in which control mechanisms are arranged in a hierarchy of five layers: the physical, vehicle regulation, platoon, link, and network layers. The architecture helps to formulate a structured approach to the design of IVHS control because:

- The hierarchy satisfactorily resolves all four dimensions of differences in control tasks.
- Each layer presents a standard, reference model to the layer above it. This provides a 'clean' interface between layers, and the design of each layer can proceed independently using the reference model of the layer below. When standardized, the reference models serve as an IVHS open systems architecture.
- Communication takes place only between adjacent layers and between peer layers. This will help to specify the communication capabilities needed to support the control system.

The focus in the second part is on a 'partially automated' IVHS system which provides information to drivers for route guidance and congestion reduction and exercises some flow control; also available are automatic cruise control and collision avoidance systems to assist drivers. However, the driver is in control and may ignore the offered advice and assistance. We show that the control tasks can be arranged in another five layer hierarchy: the physical, regulation, planning, link and network layers. The five layers are functionally similar to the corresponding layers in the first architecture. However, the reference models are significantly different.

The functions of the corresponding layers in the two architectures and the associated information requirements are sufficiently similar so that we strongly urge that future work aimed at successive refinements of either architecture should insist on a graceful migration to the other architecture. It appears at this time that the partial automation will be developed and deployed before full automation. So our suggestion is that 'partially automated' systems should be designed to accommodate extensions to the additional features envisaged in a 'fully automated' IVHS system. This is all the more important because, in our judgement, the capacity increases achievable by partial automation are likely to be very small compared with full automation and insufficient to meet the problems of congestion and safety posed by increased demand.

The basic motivation for this work is to invite discussion on IVHS architectures from relevant participants including transportation agencies, automobile manufacturers, control and communications equipment developers, and the research community. We have deliberately sketched an idealized portrait of full automation and suppressed the important concerns of system evolution in the expectation that this will help sharpen the discussion.

1 Introduction

Engineers face new challenges in the design and operation of highway systems that meet the growing demand for automobile travel. On the one hand the objective of minimizing adverse environmental impacts has been added to the continuing and pressing aim of reducing delays and congestion without compromising safety. On the other hand, the traditionally most important approach to increasing capacity by building more lane-miles of highway is often unavailable either because of prohibitive cost or public opposition.

There is a growing consensus that an adequate response to some of these challenges will take the form of an Intelligent Vehicle/Highway System or IVHS. Underlying this consensus is the opinion that an appropriate combination of control and communication technologies placed on the vehicle and on the highway can lead to significant increases in capacity and safety without requiring more land for new highways.¹ In a phrase, the idea behind IVHS is to trade off land for capital in the form of automation.² Indeed, it can be argued that automation can increase the productivity of existing capital investments in highways and vehicles.

Beliefs vary as to the specific shape of IVHS 'automation'. Clearly there is room for a spectrum of technologies. At the easy-to-implement, low-tech end of the spectrum are systems which collect, calculate, and communicate to the individual drivers relevant information about the state of the network and advice about their speed and route, and which exercise a limited degree of direct control through ramp metering. We call such a system '*partially automated*'. At the speculative, high-tech end of the spectrum are '*fully automated*' IVHS systems. The key difference between partial and full automation is that in the former vehicles are manually controlled, while in the latter they are under automatic control. Proponents of the latter believe that automatically controlled vehicles can be much more closely spaced with adequate safety and without reducing speed, leading to highway capacity increases by as much as a factor of three or four. This potentially very large payoff is the reason for the considerable interest and enthusiasm for full automation. Of course, the lack of experience in integrating the variety of technologies that will constitute a fully automated IVHS is sufficient ground for healthy skepticism.

The paper sketches both fully and partially automated system architectures, with a distinct emphasis on full automation. The first part of the paper, §2-6, is devoted to fully automated IVHS. §2 presents a rough conception of the way traffic will be organized under full automation. The control tasks that must be performed to implement this organization follow readily from this conception. In §3 we describe those tasks, and differentiate them along the dimensions of function, time scale, spatial scope and information span. In §4 we outline a five layer architecture and partition the various tasks among these layers. In §5 we propose a reference model for each layer. The reference model will help decouple the problem of design of each layer. More importantly, once the reference models are standardized and

¹The objective of minimizing adverse environmental impact is only tangentially addressed by IVHS. That objective is better served by complementary work on alternative engines and fuels, and by managing demand.

²At this time we have no quantitative estimates of these tradeoffs. Indeed a major objective of work in IVHS is to arrive at such estimates.

accepted, they constitute an *open* architecture for IVHS systems. The standards define the interface between layers and different (possibly proprietary) implementations of each layer can interwork so long as they meet the interface standards. In §6 we give an example to show how the architecture has implications for the communication needs that must be supported.

In the second part of the paper we consider partial automation for which another five layer architecture is presented in §7. The corresponding layers in the two architectures are functionally similar. Therefore we argue that further work on either architecture should accommodate migration to the other architecture. Although functionally similar, the performance of the two IVHS systems is likely to be quite different and we suggest in §8 that the capacity increases resulting from partial automation are insufficient to meet the challenges of increased traffic demand.

We conclude with some observations on the technologies needed to implement the proposed architectures. We remark that many of these technologies are already available. We call for work on systems that integrate these technologies to provide the functionalities posed in the architectures.

The proposed architectures assume ‘normal’ operating conditions. Under such conditions the overall objective of IVHS is productivity maximization, i.e., meeting the demand (measured, for instance, in vehicle-miles) at the minimum resource cost (including time, fuel, pollution). The proposed architecture is insufficient for ‘emergency’ conditions when the objective is to minimize the risk of damage rather than to maximize productivity. We do not discuss emergency control procedures, nor do we discuss how such procedures should articulate with the normal operating procedures. These issues will be discussed in a future paper. We believe, however, that our proposal can be extended to handle emergency conditions although the control procedures will be different. More significantly, considerations of failures will almost certainly place constraints on the *physical* configuration of an IVHS system. Those issues are also beyond the scope of this paper.³

The basic motivation for this work is to invite discussion on IVHS architectures from relevant participants including transportation agencies, automobile manufacturers, control and communications equipment developers, and the research community. We have deliberately sketched an idealized portrait of full automation and suppressed the important concerns of system evolution in the expectation that this will help sharpen the discussion.

2 Organization of traffic

We imagine a homogeneous automated network of interconnected highways. Appropriately equipped vehicles may enter and exit this network at various *gates*⁴ and travel through this network under automatic control. The network is embedded in a larger transportation

³See [1] for a provocative discussion of critical safety concerns in fully automated IVHS.

⁴Keywords are italicized the first time they are used.

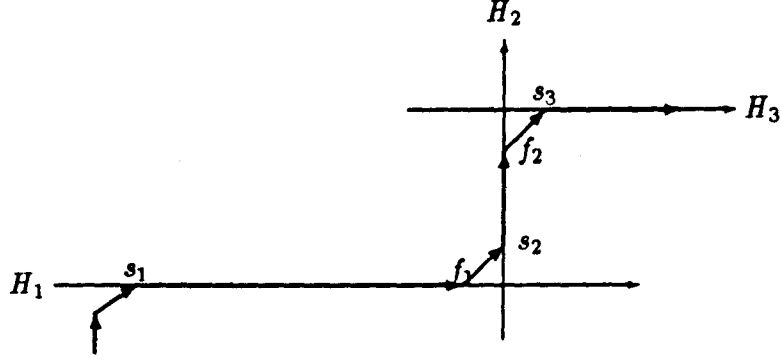


Figure 1: A route is a sequence of segments

system containing several inhomogeneous networks.⁵ For example, a vehicle leaving our network may enter an unautomated network of urban arterials. We describe how a vehicle's trip through the network from entry to exit is selected and realized.

Upon entering a gate, the vehicle announces its ultimate *destination*. The (IVHS) system responds by assigning to it a nominal *route* through the network. A route R is a sequence like

$$R = (H_1, s_1, f_1), (H_2, s_2, f_2) \dots$$

The interpretation is that the route consists of a sequence of *segments*. The first segment of the route is on the highway named H_1 starting at gate s_1 and finishing at gate f_1 ; the second segment is on highway H_2 from gate s_2 to f_2 ; and so on. It is understood that an interchange lane connects f_1 on H_1 to s_2 on H_2 , etc., as illustrated in Figure 1.

We assume that whenever a car is on an automated highway it continuously senses the *section* on which it is traveling. A section is denoted by triple (H, l, d) where H is the highway name, l is the lane number, and d is the lane section number. A section may be between 50 m to 500 m long. It is used for flow and congestion control discussed in §2. The system continuously announces a target speed $v(H, l, d)$ which cars in that section must try to maintain.

Suppose our car enters through gate s_1 at section (H_1, l_1, d_1) . The car then announces that it will exit at gate f_1 . In response, the system assigns to it a *path* (l_2, d_2, l_3) . The interpretation is that the car must change to lane l_2 , travel along it until section d_2 , and then change to lane l_3 from which it must exit at gate f_2 . See Figure 2. The path assignment procedure must balance the utilization of all lanes and exit gates.⁶

⁵The automated network may consist of segregated lanes within each highway section similar to HOV lanes. Procedures would be in place to verify that vehicles entering the automated network are properly equipped.

⁶Figure 2 is drawn to illustrate the maneuvers described below. It is not meant to suggest that entry occurs in the rightmost lane and exit in the leftmost lane. Most likely, entry and exit both occur on the right (lane l_1) and lane l_2 would be assigned according to the distance the vehicle will travel: vehicles traveling a longer distance on any highway would be assigned to lanes further to the left.

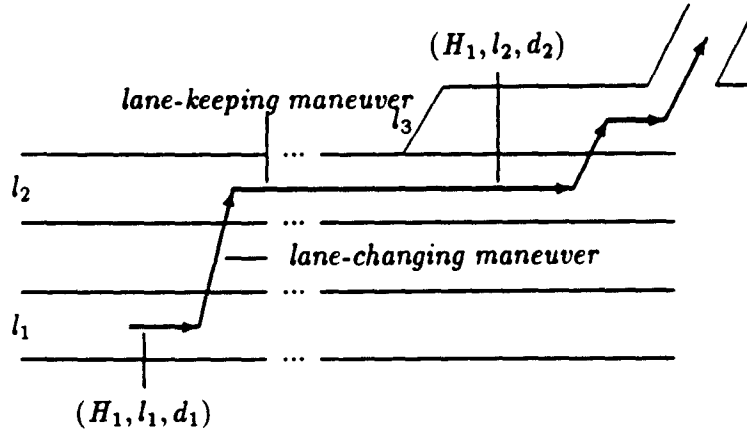


Figure 2: A path within a segment

Having received its assigned path, the car must execute an actual *trajectory* conforming to this path. A trajectory consists of a sequence of *maneuvers*:

1. An initial sequence of *lane-changing maneuvers* during which in a relatively short distance the car moves from its current lane l_1 to its assigned lane l_2 . After it enters lane l_2 , the car will execute
2. A *lane-keeping maneuver* during which it will keep to its assigned lane until it reaches section d_2 . At this point it will execute
3. A terminating sequence of *lane-changing maneuvers* to move from lane l_2 to lane l_3 which leads to the exit gate f_2 .

In order to describe how our car executes these three maneuvers we must introduce the notion of *platoon*. Every car on the highway is a member of some platoon of cars of size bigger or equal to one. A one-car platoon is called a *free agent*. The first car in a platoon of size at least two is called its *leader*, the other cars are *followers*. Only free agents may execute a lane-changing maneuver; all platoons may execute lane-keeping maneuvers. Figure 3 illustrates the definition of a platoon. The reason for this platooning strategy will become clear when we discuss lane-keeping.

We now return to our car's three maneuvers. At the beginning of its path (Figure 2) it is a free agent in lane l_1 . It executes a sequence of lane-changing maneuvers until it ends up in lane l_2 . Each lane-changing maneuver accomplishes a single lane change, so that from lane l_1 the car can move to lane $l_1 \pm 1$. While the admittedly difficult 'details' of this maneuver are yet to be worked out, the following rough description of a 'negotiate and execute' *two-phase procedure* does not seem too far-fetched. Suppose $l_2 > l_1$ so our car wants to move to $l_1 + 1$. In the first phase our free agent exchanges messages with its 'neighboring' platoon leaders in lanes $l_1 + 1$ and $l_1 + 2$ and negotiates with them the

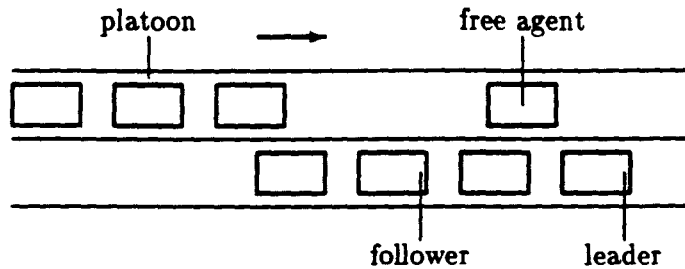


Figure 3: Definition of platoon

availability of empty space in lane $l_1 + 1$ that it needs to change lanes.⁷ Once a commitment to keep the empty space available has been obtained from the relevant neighbors, our free agent enters the second phase. In that phase it issues a command to the vehicle's feedback control system to *steer* the car into the free space in lane $l_1 + 1$. At the end of the steering control, the neighbors are informed that the procedure is complete and they are released from the prior commitment. After $|l_2 - l_1|$ such lane-changing maneuvers, our free agent is in lane l_2 and begins executing the lane-keeping maneuver. The 'negotiate' phase can be done directly among vehicles as suggested here, or it can be done in a more centralized manner. In the latter case, information about the location of each vehicle in a particular section would be maintained in a database for that section and a free agent engaged in a lane-changing maneuver would 'reserve' a space in the adjacent lane through a 'negotiate' transaction with that database. Generally speaking, in the proposed architecture decisions are made in as decentralized a manner as possible to reduce information flows and to make the system more robust.

We now describe the lane-keeping maneuver. Since cars spend most of their time and distance during a trip in a lane-keeping maneuver, this maneuver must be designed to maximize productivity, i.e., to maximize the flow in a lane. This is achieved by moving cars in platoons, see [3, 4]. Within a platoon all cars except the leader – that is, the followers – are under continuous feedback control which maintains a very close spacing (about 1 m) with the car in front. The cited references show that the steady state flow increases significantly with platoon size. With platoons of size 20 the capacity can increase by a factor of three or four.⁸ This is the reason for the platooning strategy. Studies suggest that a feedback control can be designed to provide an adequate performance in platoons of size up to 20 vehicles, see [5]. The leader must maintain the announced target speed $v(H, l, d)$ while keeping a safe distance between itself and the platoon in front of it.

A platoon leader in the lane-keeping maneuver may find that its platoon size is different from the target size (of say 20). If the size is smaller than 20 the leader initiates a platoon *merge move*; if the size is larger than 20 it initiates a platoon *split move*. The

⁷The exchange of these messages has been called a *protocolin* [2]. That reference shows how such protocols are designed and validated.

⁸Such increases could be achieved by 20-car platoons with a between-platoon headway of 60 m and a within-platoon spacing of 1 m. These numbers are derived from unverified traffic flow models, [4]. If one requires that a platoon with a braking rate of 0.3g avoid colliding with the platoon in front that is braking at 1.0g, starting at an initial speed of 50 km/h, the between platoon headway should be 40 m [3].

merge move can be implemented by another negotiate and execute two-phase procedure. In the first phase, the leader determines if the sum of its own size and that of the platoon in front of it is smaller than 20. If it is, it requests permission to merge. If that permission is obtained (which would happen unless the platoon in front was itself engaged in a merge or split move), the leader enters the second phase. In that phase it will speed up until it is sufficiently close to the tail of the platoon in front, and it changes its role to a follower so that its feedback control system will henceforth control spacing. If the permission to merge is denied by the platoon in front, the leader will wait for a fixed time after which it will try again. In a split move the leader decides to split the platoon. This is accomplished by another two-phase procedure. In the first phase the leader requests the relevant follower to change its role to a leader. When the request has been granted, the second phase is begun. In that phase, the relevant followers are informed as to who their new leader is, and thus a new platoon is formed. The new leader will slow down initially since it is too close to the platoon in front and after a safe distance is reached it will try to maintain the target speed $v(H, l, d)$ announced by the system.⁹

When our free agent enters its lane-keeping maneuver it notes that its platoon size of one is below the target of 20 and it initiates a merge move. If it is possible to do so, it will end up in a larger platoon. Its role may change from leader to follower several times while it is in lane l_2 and until it reaches section d_2 (see Figure 2) when it must begin lane-changing maneuvers in order to move to lane l_3 . Recall that in order to execute these maneuvers, it must be a free agent, which is unlikely. If it is a leader, it initiates a split move, making its immediate follower a leader; at the end of this move our car will become a free agent. If it is a follower, it requests its own leader to initiate an appropriate split move at the end of which our car will become a leader. Another split move will make it a free agent.¹⁰ Once it has become a free agent, our car enters a sequence of lane-changing maneuvers until it is in lane l_2 from which it exits at gate f_2 .

This completes our description of the way traffic might be organized in a fully automated IVHS. We now focus on the key control tasks that must be performed to implement this organization.

3 Control tasks

We describe four key control tasks beginning with decisions that have the widest impact and ending with the feedback control of a single vehicle. We give a brief account of each task and then specify its objective, time scale, spatial scope, and information span. Comments on each task are collected under "Remarks".

⁹Several aspects of this discussion need elaboration. In particular, how far must a vehicle separate itself from its neighbors to become a 'free agent' for performing lane changes, merge or exit maneuvers? These transitions must be designed on the basis of vehicle dynamics, information available, and safety.

¹⁰See [2] for a description of protocols corresponding to the split and merge moves.

3.1 Route and flow control

When a car starts its trip at some gate (H, s) and it announces its destination gate (J, f) , the system assigns to it a route

$$R = (H_1, s_1, f_1), \dots, (H_n, s_n, f_n)$$

where $(H_1, s_1) = (H, s)$ and $(H_n, f_n) = (J, f)$. Thus this decision is a function:

$$(\text{origin, destination}) \rightarrow \text{Route} \quad (1)$$

This function could be implemented as a routing table. The route selection is made jointly with the selection of equilibrium speeds in each highway section, i.e., with the selection of a function

$$(H, l, d) \rightarrow v_e(H, l, d) \quad (2)$$

Objective. The routes and section speeds must be selected so that the equilibrium or steady state traffic flows in the system are optimized according to some criteria. Possible criteria include: minimizing total trip time across the network, minimizing the time of each route at the moment the request is received, and minimizing the maximum or a high percentile of excess travel time (relative to the free-flow travel time).

Time scale. The maps (1), (2) are updated only when there is a significant and long term shift in demand (e.g. morning rush hour, afternoon, evening rush hour, night; weekday, weekend, etc.) or in network capacity (e.g. certain lanes may be closed, driving conditions on some highway segments may have changed due to weather, etc.). The frequency of such changes is on the order of once per hour on the average, but these changes could occur rapidly in response to incidents.

Spatial scope. Changes in the routes and equilibrium speeds affect traffic throughout the network.

Information span. Any procedure that will accomplish this task satisfactorily will require systemwide information on demand, capacity, and possibly of existing flows. This information will be at a macroscopic level.¹¹

Remarks. This description implies that route selection is done at a centralized, systemwide level according to systemwide criteria. Alternatively, in a more decentralized implementation, a computer on board the vehicle could select its own route based on traffic conditions broadcast by the system. In the latter case each route would presumably be selected to minimize individual trip time which may compromise systemwide performance criteria such as time for all trips. The point is that the proposed scheme can accommodate many possible implementations.

¹¹By macroscopic level we mean that traffic is treated as a compressible fluid rather than as a collection of individual vehicles. The fluid is described by macroscopic variables such as density, speed, and flow.

3.2 Path and congestion control

When a car enters a highway H at a gate s in section (H, l, d) and announces that it will exit at gate f , the IVHS responds with a path assignment (l_2, d_2, l_3) . d_2, l_3 are determined entirely by the exit gate f (see Figure 2), and that information can be stored in a permanent table. Thus the only decision to be made in path assignment is the selection of the lane l_2 over which the car will travel most of the time it is on this highway. If H has several lanes, efficiency dictates that the traffic should be evenly distributed across lanes.¹² This is a fairly trivial task: for example, assignment of cars to lanes in ‘round-robin’ order will achieve a balanced distribution of traffic across lanes. The ‘round-robin’ rule could be modified so that vehicles that need to exit “soon” are in the “near” lane, while those traveling longer distances could better justify the lane-changing overhead incurred in moving to the “far” lane.

Path assignments may become more complicated than implied above when the system must respond to accidents or other blockages. It may become necessary to change a route after a vehicle has entered a section if the originally assigned route suffers a problem.

The system must also announce the platoon target speed $v(H, l, d)$ in each section of the highway. This speed will normally be the same as the equilibrium speed $v_e(H, l, d)$ of (2). However, in order to reduce congestion the target speed may temporarily deviate from the equilibrium speed. We discuss this briefly below. Lastly, the system must announce the target platoon size $p(H, l, d)$ for each section of the highway. This size depends on the equilibrium flow on the highway: the size increases with the flow [4]. This is a trivial task and we will not discuss it further.

Small random fluctuations in the flow can create a decrease in the platoon headway in a section. As a result platoons in that section may be forced to slow down. This causes decreases in headways further upstream following which those platoons may slow down. The congestion could ‘travel’ upstream in a wave and could persist for a considerable period causing significant, sometimes severe, decline in traffic flow. However, if the onset of disturbance is detected, and the target speeds are appropriately modified, then the disturbance can be eliminated quickly with negligible reduction in flow. A model for congestion control, presented in [4], suggests how automation can reduce fluctuations in the flow and an asymptotically stable longitudinal control system can damp these out without creating a traveling wave.

Objective. Paths must be selected to balance traffic across lanes to prevent inefficient utilization of lanes. The section target speeds $v(H, l, d)$ must be selected to minimize congestion while following the equilibrium speed on the average.

Time scale. In the absence of real data our guess is that disturbances leading to congestion may occur every few minutes in each lane-mile during heavy traffic. To reduce congestion quickly once it occurs, the target speed should be updated a few times each minute, [4]. However, this estimate is based on an unverified model. The target speed

¹²Efficiency may also be increased by specializing lanes by vehicle type, e.g. trucks, buses, automobiles.

update could even be on a second-by-second basis when needed to adapt to rapidly changing conditions, although of course in general it would be better for passenger comfort and energy efficiency not to change speeds too often.

Spatial scope. Path assignment only affects one car at a time. Congestion control modifies section speeds up to about 1,000 m from the place that is congested.

Information span. A round-robin path assignment scheme requires no information. Congestion control in any section requires knowledge of speeds and headways within 1,000 m on either side of the section. This information can be at a macroscopic level, and may be obtained by ‘filtering’ data on car counts and passing speeds [6], or by data obtained from platoons.

Remarks. Simple heuristics can be incorporated in procedures to be used in ‘path assignment’ algorithms. Such algorithms would be linked to algorithms that estimate congestion, and detect and respond to incidents.

Random fluctuations in traffic flow now arise from non-uniformity and randomness in driver behavior and random arrivals and departures of vehicles to and from the highway. The first type of randomness would be reduced by a partially automated IVHS system which announces advisory section target speeds [4, 7]; the reduction would be much greater with full automation since platoons would automatically track target speeds (whereas drivers may be less responsive or may even ignore advisory speeds). The second source of randomness can be partially controlled by ramp metering. The increased control over traffic flow obtained by automation should reduce randomness and congestion. However, quantitative estimates of these reductions must await appropriate simulation experiments.

3.3 Platoon maneuvers

A platoon can execute lane-changing maneuvers (if it is a free agent) and lane-keeping maneuvers. A lane-changing maneuver involves a two-phase procedure: a negotiating protocol is executed to get a commitment of free space from neighboring platoons, followed by issuing a steering command. A lane-keeping maneuver is executed by the platoon leader. Most of the time the leader tracks the announced target speed $v(H, l, d)$. The leader may also engage in platoon merge and split moves in order to maintain platoon sizes close to the announced target size and to accommodate lane-changing maneuvers. Lane-keeping maneuvers are also implemented by ‘negotiate and execute’ two-phase procedures.

Objective. These maneuvers must be executed safely. The negotiating protocols must be correct; they must take into account possible error conditions. For example, errors can occur during communication. Each car must keep track of its platoon data. This data must be updated whenever it changes.¹³

¹³[2] gives a list of data that is adequate for the protocols considered there.

Time scale. A free agent will engage in lane-changing maneuvers very infrequently (perhaps twice per segment). The merge and split moves should also be infrequent. The time between such maneuvers should be on the order of minutes.

Spatial scope. Platoon maneuvers only affect neighboring platoons.

Information span. Platoon leaders need the target speed $v(H, l, d)$ and the target platoon size $p(H, l, d)$. They should be able to communicate with neighboring platoons. They should be able to set feedback control laws on their own vehicles.

Remarks. Automation of platoon maneuvers and automatic control of vehicles (§3.4) are the two distinguishing features of full automation. They constitute a radical departure from current practice, and their implementation will require a major research and development effort. The negotiation protocols are similar in structure to computer-communication protocols [8]. They can be described and verified for correctness using high-level description languages (see, eg. [9]). A version of the platoon negotiation protocols is presented in [2]. Implementation of the negotiating protocols requires an underlying communications network to exchange protocol messages. Some remarks on possible communications technologies can be found in §8.

3.4 Feedback control

The ‘execution’ phase of a platoon maneuver is implemented by vehicles under automatic feedback control. Three kinds of feedback control laws are needed. A car which is a follower must maintain a constant spacing with the car in front and it must stay in the center of its lane. We call this *spacing control*.¹⁴ The car which is a leader executing a lane-keeping maneuver is under *tracking control*: it must track the target speed while keeping at a safe distance from the platoon in front of it. If the leader is engaged in a merge move it tracks a speed somewhat larger than the target speed. When a free agent executes the second phase of its lane-changing procedure, the car is under *steering control*: it must move from its current position to a specific point in an adjacent lane within a certain amount of time. This would also happen when merging into or out of a main traffic flow.

Objective. The three kinds of feedback control laws must maintain stability while accomplishing the spacing, tracking and steering tasks.

Time scale. The time scale is on the order of 0.3 s, based on time constants of dynamic models of contemporary cars.

Spatial scope. The direct effect of feedback control is limited to one car.

Information span. This is still unclear since control designs have not been worked out. For example, the longitudinal control design of [5] requires that the controller know its own

¹⁴It is customary to separate this into *longitudinal* and *lateral control* [5].

car's state (position, velocity), the spacing between its car and the car in front, and the velocity and acceleration of the leader and the car in front. In addition, controllers may need to 'preview' the shape of the highway in terms of grade and curvature. While the information needed is certainly limited to a short distance from the position of the car, sophisticated sensors may be needed to obtain it, [10], although some very simple schemes may suffice, [11].

Remarks. The design of feedback control laws, the difficulty of implementing them, and their performance depend critically upon the information and the actuators available. The information in turn is a function of (1) the sensors on board the vehicle for measuring velocity, acceleration, and relative distance and speed between adjacent vehicles, (2) the sensor data from one vehicle communicated to neighboring vehicles, and (3) information about the geometry (curvature and gradient) of the highway communicated to the vehicle. (This last information would be used in a 'feedforward loop' to enhance stability and smoothness of the ride.) The response characteristics for braking, acceleration, and steering actuators limit the overall performance. Generally speaking, the theory and practice of the design of control laws seems to be sufficiently well advanced so that one can confidently predict the creation of satisfactory designs for the three feedback control tasks listed above. On the other hand, suitable sensors and actuators are still under development, although it is likely that significant advances have been made in industrial laboratories but are not reported in the published literature.

4 IVHS control architecture

Figure 4 gives a block diagram description of a five layer control architecture. Starting at the top of the hierarchy the layers are named: network, link, platoon, regulation, and physical layers. The functions of the top four layers correspond in order to the four tasks outlined in §3, namely: route and flow control, path and congestion control, platoon maneuvers, and feedback control. The function of the physical layer is to provide the regulation layer with relevant sensor data and to receive from it the acceleration, braking, and steering actuator signals.

The hierarchy resolves the four dimensions of difference between tasks. Each task is lodged in a separate layer; the frequency of decisions increases, the spatial scope reduces, and the information span is more localized as one moves down the hierarchy.

It seems natural, as suggested in the figure, to distribute the control task in each layer among several identical controllers. For example, there would be one controller per vehicle at layer 4, one controller per platoon at layer 2 (which would be distributed among the vehicles in the platoon), one controller at layer 3 for each highway *link*¹⁵, and one or a few controllers at layer 1. With that distribution of control tasks, a controller at each layer 'supervises' one or more controllers at the layer below.

We note one feature of this supervision. At the regulation layer, at any time, a

¹⁵A highway link would consist of several sections; it would be a few miles long.

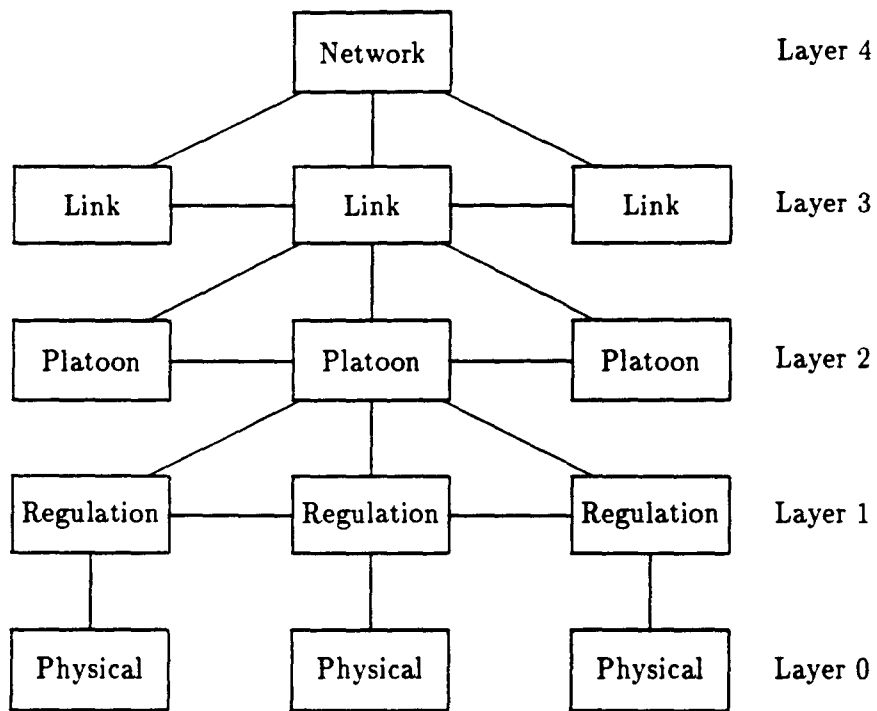


Figure 4: IVHS control architecture

controller is implementing one of the three control laws discussed in §3.4. The controller continuously changes the actuator commands in response to changes in sensor signals. On the other hand, it switches control laws when so ordered by its platoon layer supervisor. Similarly, at any time, a platoon layer controller is engaged in a maneuver. During a lane-keeping maneuver a leader changes its target speed when so ordered by its link layer supervisor. A link layer controller computes a new target speed $v(H, l, d)$ or a new target platoon size $p(H, l, d)$ either in response to a detection of a disturbance or because it receives a new equilibrium speed from its supervisor at the network layer. Thus decisions at each layer are changed either in response to information from ‘below’ or because of new commands received from ‘above’. Some aspects of such hierarchical control systems are discussed in [12, 13].

It is important to observe that a controller at each layer needs to communicate only with controllers at adjacent layers and its peers.¹⁶ More interesting is the fact that a controller at each layer receives ‘state information’ from controllers below it and returns ‘commands’ to those controllers. Thus commands flow down the hierarchy and information flows up. This has the important implication that each layer presents to the layer above it a model of the system, and the controller at any layer can be designed in terms of this model alone. We explore this in §5. The structure of the information flows also has implications for the communications infrastructure needed to support the control architecture. This is explored in §6.

This architecture design is based on the principle that a decision must be taken close to a point where the information on which it is based is available. Thus, for example, actuator signals must be calculated by a controller in each vehicle, even though they could be calculated remotely. It is believed that this increases robustness of the design, localizes the impact of future changes, and reduces communication costs and delays.

We repeat one point noted at the end of §1. This architecture is designed for ‘normal’ operating conditions where the aim is to maximize productivity. It is not intended to address emergency conditions. When such conditions occur as during an accident or malfunction, in order to minimize response time action must be initiated by a controller ‘close’ to the point where the emergency is detected. It is possible that this requires a separate set of emergency control rules which ‘override’ the normal rules. The single objective of these rules is to minimize the risk of damage.

5 Reference models and open architecture

Perhaps the most important advantage of the proposed architecture is that it permits a separate specification of the control task at each layer and the controller design at one layer can be carried out independently of the designs at other layers. The controller design at a layer is done in terms of a reference model of the layer below. We briefly describe the four reference layers, beginning with the link layer and proceeding down the hierarchy of Figure 4.

¹⁶Controllers at the same layer are peers.

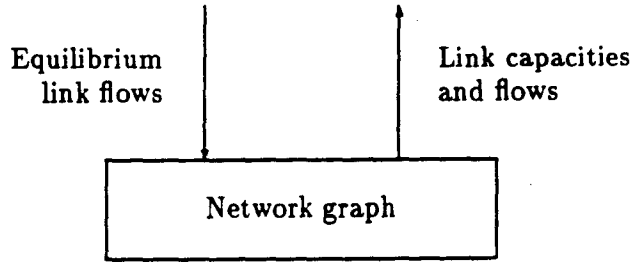


Figure 5: Link layer reference model

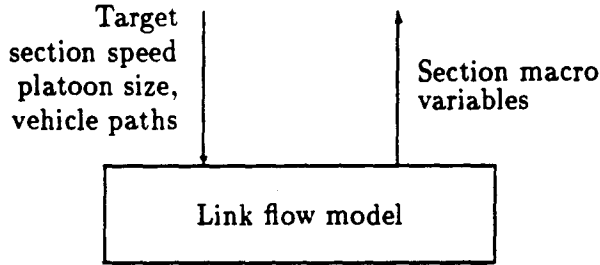


Figure 6: Platoon layer reference model

The **link layer reference model** is quite simple. It provides an aggregate description of the flows in each link. It also indicates to the network layer any significant capacity changes that may have occurred in the link. Thus this reference model can be presented as a graph each of whose links is characterized by the current capacity and the current flow. Such a reference model suggests formulating the layer 5 route and flow control task as a mathematical programming problem of network flow optimization. See Figure 5.

The **platoon layer reference model** is a dynamic model describing the behavior of platoons on a link.¹⁷ The model equations involve macroscopic variables such as average speed, density and flow in each segment. The ‘inputs’ to the model are the section target speed $v(H, l, d)$ and platoon size $p(H, l, d)$ periodically updated by the link layer, and the path assigned to each vehicle. Its ‘outputs’ are the macroscopic variables (speed and density) over each section in the link. (The outputs may be computed from data provided by each platoon.) The congestion control policy of the link layer is based on this reference model. Examples of such models may be found in [7, 4]. See Figure 6.

¹⁷Recall that a link is about a mile long; it is made up of sections between 50 m and 500 m long.

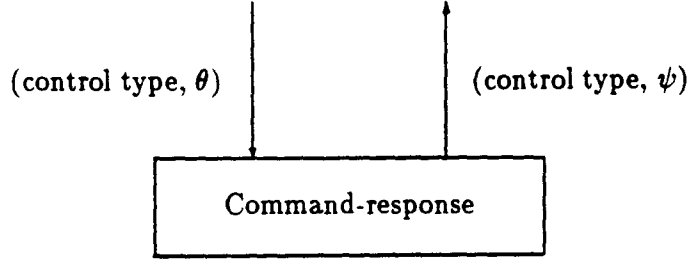


Figure 7: Regulation layer reference model

The **regulation layer reference model** is a ‘command-response’ model of the feedback controlled dynamics. The commands are issued by the platoon layer. Those commands can be encoded as parameters for each of the three types of control laws. For example, spacing control is parametrized by the space that a follower must maintain with the car in front of it. A tracking control is parametrized by the target speed that a platoon leader must maintain and the safe distance that it must keep from the platoon in front of it. Lastly, steering control is parametrized by the point on the adjacent lane to which a free agent must move and the time within which it must reach that point. The response to each parametrized command can itself be in the form of parameters that indicate successful execution of the command or the occurrence of some error condition.¹⁸ Thus the regulation layer reference model takes the form:

$$(\text{control type}, \theta) \rightarrow (\text{control type}, \psi)$$

where ‘control type’ is ‘spacing’, ‘tracking’, or ‘steering’, and θ is the associated parameter vector; and ψ is the parameter vector of the response, see Figure 7.

The **physical layer reference model** is given by a differential equation model of the vehicle, actuator, and sensor dynamics.¹⁹ The three sets of control laws (spacing, tracking, and steering control) are designed using this reference model, see Figure 8.

Each layer’s reference models summarize how the system appears to the layer *above* it. There are also *peer* reference models that are used to design peer interaction. We briefly discuss these. Exchange of information among neighboring link layer controllers is needed in order to maintain continuity from one link to the next, and to warn of changes in flow or capacity. Such exchanges would be governed by a **link layer peer reference model**.

In a lane-changing maneuver a free agent carries out a ‘negotiate and execute’ two-phase procedure. In the first or negotiation phase, the free agent exchanges a structured set of messages – a protocol – with neighboring platoons at the end of which the free agent has obtained a commitment of empty space in the adjacent lane. The correctness of this

¹⁸Error conditions would indicate to the platoon layer supervisor why the command cannot be executed or how the result of the execution differs from the expected.

¹⁹There will be several models depending on vehicle, actuator and sensor type.

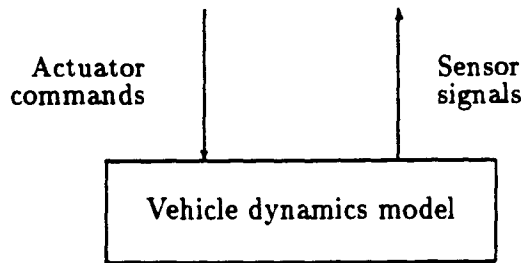


Figure 8: Physical layer reference model

protocol is determined within the context of a **platoon layer peer reference model**. A useful formalism for such a model is provided by finite state machines, see [2]. Such finite state models serve the same function as the differential equation models of the physical layer reference model. The only difference is that the states of the finite state machine represent discrete logical stages in the execution of the protocol rather than the continuously varying states of the vehicle model.

Suppose the regulation layer is calculating one of its three types of feedback control, for instance, the spacing control law. The calculation may depend on certain parameters of the other vehicles in the platoon. For example, the maximum permissible acceleration or braking may be limited by the vehicle in the platoon with the least performance. Thus, the regulation layer controller of a vehicle will need to exchange messages with its peers in the same platoon to determine the necessary parameters. Such an exchange would be governed by a **regulation layer peer reference model**.

The phrase '*open systems architecture*' has gained wide currency from its adoption and use in the context of computer-communication networks. The multi-layered architecture of these networks is standardized by bodies representing equipment manufacturers, users, service providers, etc. The standards describe reference models (i.e. the interfaces between layers and between peers) and the corresponding protocols. The term 'open' emphasizes the fact that equipment which conforms to the standards can interwork with any other conforming equipment, even though the equipment designs may be proprietary and different. The rapid advances in data networks are in part due to the industrywide adoption of open system architectures.

Similar rapid advances are to be expected in automated IVHS technologies once the relevant participants (vehicle manufacturers, highway authorities, communication and control equipment suppliers, and representatives of the driving public) adopt an open IVHS architecture. A tremendous amount of work will be needed to establish such standards because the issues are complex, the stakes are enormous, and the impact will be long-lasting. However, without such standards equipment on different vehicles and on different

highways will be incompatible, much more expensive and far less productive.²⁰

6 Communication requirements

We have seen that in order to carry out its task, a controller will need data from its peers, from controllers in the layer below it, and from vehicles and highway sensors. The controllers will also need to communicate decision variables (target speed, actuator controls, etc.) to subordinate controllers. The ‘time scale’ of each task determines the frequency with which certain communication of data and decisions will be repeated. Once the controllers and their data requirements are sufficiently specified, we can obtain the average bandwidth requirement (measured in bits per second or bps) between every pair of nodes.²¹ Specifying the requirements in terms of average bps may be too crude, and one might also specify other ‘quality of service’ parameters such as peak bandwidth, error rate and delay. We call these **end-to-end bandwidth requirements**.

An IVHS communication network will have to meet these end-to-end bandwidth requirements. As a rough approximation we can think of a communication network as a graph with the nodes as specified above and with some pairs of nodes connected by communication links of certain capacity. To send data from one source node to another sink node, that data may have to be routed over several communication links depending on the graph of the communication network. It is then relatively straightforward to check whether the network meets the end-to-end bandwidth requirements.²² It is more work to figure out whether requirements on delay, errors, etc. are met.

Many different networks can meet a given set of requirements. An important part of IVHS communication network design will be to articulate and analyze several options. We conclude this section with a small example.

Suppose the feedback law for longitudinal control requires that the acceleration u of a follower in a platoon be set according to

$$u = f(\Delta, s, v, \dot{v})$$

where Δ is the distance between the follower and the car in front of it, s is the speed of that car, and $v, \dot{v}$ are the speed and acceleration of the leader of the platoon. Suppose Δ is sensed directly by the follower, s is sensed by the car in front, and $v, \dot{v}$ are sensed by the leader. Suppose each number is represented by two bytes. Suppose finally that u must be evaluated N times each second, and that there are n cars in the platoon. Then we can calculate the end-to-end bandwidth requirements as follows.

Label the cars $1, \dots, n$ with car 1 as the leader. Then we must have

²⁰Firms seeking short-term advantage have a predictable tendency to resist the compromises needed to arrive at standards. However, as the experience of the computer and communications industries shows, the adoption of standards dramatically increases market size to the benefit of all.

²¹A node is any source or sink of data. Examples of nodes are controllers, sensors, and passengers.

²²By ‘straightforward’ we mean only that one can use several known techniques; we do not mean that the calculations involved are easy.

1. $1 \rightarrow i, i = 2, \dots, n$ 4-byte messages for $v, \dot{v}$ from the leader to each follower N times each second;
2. $i \rightarrow i + 1, i = 1, \dots, n - 1$ 2-byte messages for s from car i to the car $i + 1$ behind it N times each second.

Suppose the communication is carried out over a single broadcast radio channel. Then item 1 above will require a single 4-byte broadcast

$$1 \rightarrow \{2, \dots, n\}$$

N times each second. Thus the channel must carry information at the rate of

$$I = 8N \times (4 + 2(n - 1)) \text{ bps}$$

Taking $N = 30$ (which corresponds to a sampling interval of 33 ms) and a platoon size of $n = 20$ gives $I \sim 10,000$ bps. This information consists of 20 separate broadcast messages (one per car) per sampling interval. Each message will require some 'overhead' bits for synchronization, error checking, etc. If we take this to be 4 bytes, we get an overhead bandwidth of

$$O = 4 \times 20 \times 30 = 2,400 \text{ bps}$$

Finally, we need to have a 'guard time' between two successive broadcast messages. Suppose this reduces the 'duty cycle' of the channel to 50 %. Then the radio channel capacity should be

$$C = 2(I + O) \sim 25,000 \text{ bps}$$

If the same channel is used by several different platoons, the capacity will be increased correspondingly. In this calculation no provision is made for acknowledgment of the messages. Such a provision may double the capacity.

7 IVHS architecture for partial automation

As in part one we start with a description of how traffic might be organized. From this we infer a set of control and management tasks which we arrange in a five layer hierarchy. Since the traffic organization, tasks and architecture are similar to the case of full automation, we will be brief.

Properly equipped cars enter and exit the partially automated network at controlled gates.²³ The entrance gate is metered, so the system may exercise some flow control. The entering vehicle announces its ultimate destination in response to which the system suggests a route through the network. A route is a sequence of segments exactly as in Figure 1.

As the car proceeds along a particular highway segment, the system provides information and advice regarding the vehicle's path along the segment. These may include:

²³Unlike full automation, partial automation can permit both equipped and unequipped vehicles, but performance will degrade if there are too many unequipped vehicles. It is an important open question to determine the fraction of vehicles that should be properly equipped to obtain significant improvement in performance.

- Announcing incidents downstream so that drivers may change lanes accordingly;
- Suggesting the speed on each section so as to reduce the propagation of congestion;
- Suggesting transfer to alternative roads in case of long-lasting congestion;
- Estimating travel times in response to driver inquiries.

The driver absorbs this information and advice and plans the vehicle's path, i.e., which lanes to occupy and what target speed to project. The actual *plan* selected will also depend on the driver's goals, predisposition, experience, and understanding of the system's operations. The plan may get revised along the way as new information is received.

The plan is executed by controlling (driving) the vehicle. In effect, the driver implements a feedback control law. In this implementation, the driver makes use of his or her own sensory data (vision, hearing, etc.), standard car sensors that indicate speed, engine rpm, etc., and signals from more sophisticated collision warning systems. The driver may also select pre-programmed feedback laws such as cruise control and collision avoidance.

This description suggests four tasks which can be arranged in the five layer hierarchy of Figure 9. As in the architecture of Figure 4, the network layer is responsible for *route and flow control*. It suggests the route to each entering vehicle. It exercises a certain degree of flow control at the entrance gates. Lastly it calculates the equilibrium speed for each section. (See (1), (2).) These decisions are based upon estimates of current traffic conditions obtained from layer 3. The decisions seek to maximize some systemwide criteria and hence can appropriately be formulated as a mathematical programming problem. However, the link layer reference model on which such calculations are based will be more complex than in §5. Under full automation, it is assumed that cars follow the routes and speeds suggested by the system. But now the model will have to include some representation of the way drivers will respond to suggested routes and speeds.²⁴

The function of the link layer is also similar to its counterpart in Figure 4, namely *path and congestion control*. It must provide information about incidents and suggest target speeds for each section based on equilibrium speeds received from the network layer and aggregate traffic data for each section received from roadside sensors or from the vehicles themselves. The selection of section target speeds is based upon some model of congestion propagation. Such planning layer reference models in turn are based upon driver response characteristics, see [14, 15, 16, 4]. Again the reference model will be more complex than in §5.²⁵

The planning layer is responsible for *path planning*. It must take information about incidents, suggested target speeds and estimated travel times, compute a plan, and forward it to the regulation layer. It must also revise plans as new information is received. In dynamic route guidance, a vehicle-based or wayside system computes an 'optimal' path based on current network conditions and gives the driver a sequence of actions (turns,

²⁴Drivers will probably ignore route advice if, in their opinion, it does not minimize their travel time.

²⁵Layers 3 and 4 comprise ATMS (Advanced Traffic Management System); layer 3 and part of layer 2 comprise ADIS (Advisory Driver Information System).

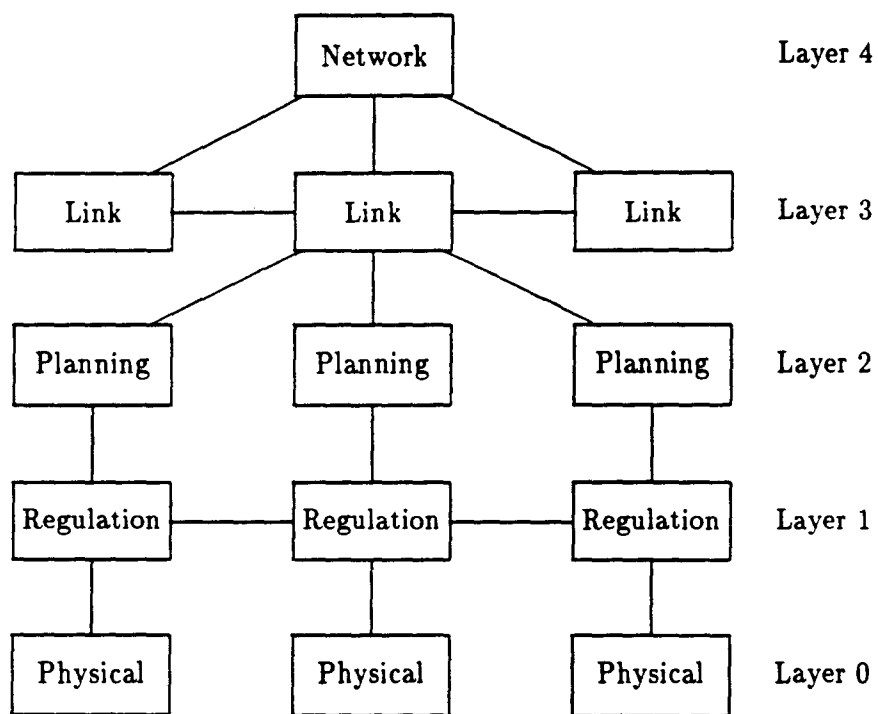


Figure 9: Partially automated IVHS control architecture

lane changes, etc.)—so path planning is automated here. Alternatively, the information is conveyed to the driver who uses it to choose a plan. Since we have no reliable understanding about how drivers plan on the basis of such information the planning layer reference model is likely to be more complex. It is to be hoped that experiments now underway [17, 18, 19, 20] will improve our understanding.

The function of the regulation layer is to implement the plan developed by the planning layer. The driver implements this using his or her own sense data, data from vehicle sensors, information from the roadside and from other vehicles. The driver may also ‘off-load’ some of the feedback control tasks to pre-programmed control laws such as automatic cruise control, automatic braking, and collision avoidance systems. If the planning layer is automated, it must be designed on the basis of some regulation layer reference model. Since most of the regulator layer will be implemented by the driver, this reference model is likely to be quite complex.

The driver and the pre-programmed laws both assume some reference model of the physical layer. The two reference models will be different. Once again we know very little about how the driver understands the dynamics of the vehicle.

In comparing the two architectures we see that the functions of the corresponding layers are quite similar and the information required at each layer is also similar; hence the resemblance of Figure 9 to Figure 4. The critical difference is that under partial automation the vehicle is under driver control and so

- the traffic flow characteristics on each highway segment are determined by the response of drivers to traffic conditions;²⁶
- the driver may accept, reject or modify the advice about route and speed given by the system.

As we argue in §8, these differences significantly reduce the performance of partially automated IVHS as compared with full automation.

8 Full or partial automation

The comparison is based on very rough estimates of the relative costs and benefits of the two systems. In comparing costs we can distinguish between costs of sensors, communications, computation, and control.

We have seen that the two architectures are similar in function and in the information requirements at each layer. Hence our guess is that the cost of sensors and communication systems in the two architectures will be comparable. The most significant difference is that the platoon layer in Figure 4 requires real time communication between neighboring vehicles and sensors that the planning layer in Figure 9 need not use.

²⁶Typically, traffic flow is characterized by an ‘equilibrium’ speed vs density curve; and drivers accelerate or decelerate depending on whether the density is smaller or larger than its equilibrium value [14, 15, 16].

The computation capabilities needed at the corresponding layers is similar. The calculations of the network and link layers under partial automation are likely to be more difficult (because the corresponding reference models are more complex) than under full automation. Even primitive versions of the planning layer will require some on-vehicle computer assistance to the driver. The incremental cost of the greater computation capacity needed by the platoon layer will be minimal.

The regulation layer under full automation will be significantly more complicated and this will require electronic actuator systems (for acceleration, braking, steering) not needed by the driver-implemented regulation layer. It is, however, quite likely that some of these systems will be installed even under partial automation in order to implement such pre-programmed control features as automatic cruise control and collision avoidance. In this case, the incremental cost of the enhanced control capability needed under full automation may not be too great.²⁷

The benefits of automation can be grouped under increases in capacity and safety. Partial automation will improve safety by providing critical information (eg. collision warning) to the driver, or by overriding driver control (eg. collision avoidance). Full automation is likely to yield greater safety because of more effective and predictable control of the vehicle. The major difference in benefits is certainly going to be in terms of capacity increase.

Under partial automation, capacity increase is achieved because travel time is reduced since drivers will have more accurate and timely information about traffic conditions and advice about the best routes. Simulation and analytical studies, and data from demonstration experiments suggest little or no improvement under recurrent congestion and some improvement under incident induced congestion.²⁸ One may with confidence suggest an upper bound of 15 % on the capacity increase from partial automation. The 'bottleneck' element in a partially automated system will continue to be, just as it is today, the driver response characteristic, i.e. the regulator layer in Figure 9.

It is precisely the elimination of this bottleneck that is promised by full automation. The regulator and platoon layers of Figure 4 can, it now appears, permit increases in capacity by as much as a factor of three to four, and the greater degree of control should reduce the propagation of recurrent congestion. It must be added immediately that these estimates are based on (1) the assumption that the technologies needed to implement full automation are feasible, and (2) simple models of traffic flow under full automation, see [4, 5, 3].

To summarize the comparison of costs and benefits:

²⁷ However, full automation may require a physically separate, hence more expensive, system, while under partial automation cars that are not properly equipped need not be excluded from the system.

²⁸ A simulation study based on the CACS project suggests that travel time in Tokyo could be reduced by 6 % [21]; U.K. researchers estimate an average benefit of 10 % from dynamic route guidance; preliminary results from the Berlin route guidance experiment show no savings in average travel time under normal conditions [18]; simulations of the Santa Monica freeway (SMART) corridor suggest insignificant savings under recurrent congestion and savings on the order of 10 minutes for a 40 minute trip under incident induced conditions [20, 22]; a careful examination of data on driver response to advisory speeds posted by the Dutch Motorway Control and Signalling System showed no increase in capacity [7]; theoretical considerations also suggest little or no benefits from route guidance under recurrent congestion [23].

- Partial and full automation will require many similar subsystems. The regulation layer system under full automation will be more expensive, but it is impossible at this time to assess the extra cost. Full automation may be significantly more costly if it alone will require a physically separate network, as seems likely at this time.
- Full automation may provide more safety because of better sensors and control, and avoidance of driver errors.
- Productivity improvements under partial automation are limited by driver response and hence unlikely to be greater than 15 %. It may be as much as 300 % under full automation.

Two conclusions urge themselves from this analysis:

- The implementation of partially automated IVHS must make sure that it can easily migrate to full automation.
- A concerted effort must be made to determine the feasibility of full automation.

9 Conclusions

We begin by addressing potential criticisms of the proposal presented here. We then outline one fruitful path for future work.

The idea of organizing a complex engineering system in a layered hierarchy is well-established. Two remarkably successful examples are the structuring of computer operating systems [24] and the open systems architecture of communication networks [8]. Thus an attempt to organize an IVHS system according to this paradigm is surely non-controversial. Of course, even within this paradigm there is plenty of room for disagreement and debate not only over details but over its most prominent features. We hope this proposal will encourage such debate.

Perhaps the most radical criticism will come from advocates of a 'low tech', partially automated IVHS in which the individual car is under manual control. The criticism can be both negative and positive. Negative criticism would charge that placing cars under automatic control will induce a major change in the way the vehicle and highway systems cooperate, requiring technologies that are either unproven or not yet developed. Another criticism, quite distinct from technological considerations, might be that drivers would be unwilling to relinquish control over their vehicles, and so full automation will never obtain public acceptance.²⁹ As a result the risk involved in pursuing the 'high tech' option of full automation is enormous.

²⁹ Another objection has to do with questions of liability in case of accidents. This is not an appropriate place to address the sociological, political and legal projections underlying these criticisms. We also leave aside another approach, which will surely gain more advocates over time, which holds that the best way to handle increased traffic demands is by mass transit or by reducing demand through congestion pricing.

Positive criticism would hold that partial automation will require minimal changes in highway technology and that most of the new technology will be put on board private vehicles. (Therefore 'public' expenditures will be much less compared with full automation.) Moreover, this new technology can be developed and deployed in independent increments. For example, cars may carry terminals which display their location on maps; they may be equipped with collision warning and avoidance systems; and drivers may access communication channels over which they could be informed of incidents and seek guidance for routes.

In essence, the criticisms amount to the judgment that upon considering the current state of the technology, the ease of deployment, and the costs of partial vs full automation, prudence dictates in favor of partial automation.

We believe that such a judgment is premature at best, and harmful at worst. In the first place, as we have shown, the commonalities between partial and full automation are so great that efforts aimed at developing a partially automated IVHS must accommodate its enhancement to full automation. In the second place, as we have suggested, the published evidence points to the conclusion that productivity increases from partial automation will be very small and may not be cost effective. On the other hand, full automation does offer the potential of very large productivity increases, although the uncertainties are very high as to whether and when this potential may be realized. It is only sensible to conduct research to reduce those uncertainties to a point where sound public decisions can be made.

We make one suggestion about future work. The basic technologies needed for IVHS—communications, sensors and actuators, control, and computation—are sufficiently advanced that effort should focus on experiments which will demonstrate the feasibility of IVHS, and which will lead to more reliable estimates of the costs and benefits of large-scale IVHS deployment.

Work on IVHS should design and develop in 'parallel' all layers of the architecture. Care must be taken to ensure that full functionality is implemented in each layer; and that all layers will work together. The latter condition will require continuous interchange among those involved in the different layers to specify the interfaces as completely as possible. The first condition may be ensured by specifying the design goals of each layer as early as possible and by insisting on meeting those goals.

In our opinion, the danger is that the overall IVHS design will be compromised as effort is diverted to pursue opportunities to develop subsystems that implement one or another function in a particular layer. It is easy to imagine, for example, that work on the regulator layer of Figure 4 will lead to subsystems that provide collision warning or collision avoidance. The temptation will be great to pursue development of such subsystems even at the cost of postponing the demonstration of the overall IVHS system. Perhaps one way of being conscious of this danger is to formulate very concrete goals that require more or less complete functionality of full automation. We suggest two such goals:

- Demonstration of a 20 vehicle platoon that can execute lane-keeping maneuvers with an average speed of 80 km/hour and a headway of 1 m.

- Demonstration on a 40 km stretch of highway of a system that can estimate the state of the traffic in each 500 m long section every 15 seconds, detect or infer incidents, compute target speeds to reduce the propagation of congestion, and communicate the target speed to every vehicle.

To achieve the first goal it will be necessary to implement layers 3 and 4 of Figure 4 in almost their full functionality.³⁰

To achieve the second goal it will be necessary to implement layer 2 of Figure 4. In our opinion, extending such an implementation to include the functionality of layer 1 is conceptually straightforward since the information needed will be available and the main remaining task will be to design and implement appropriate numerical algorithms to calculate the routing tables and the equilibrium speeds.

The word 'sketch' in the title needs emphasis and elaboration. To judge a proposed architecture we need to assess how much it will help in an IVHS control design. At a minimum, the architecture merely serves as a convenient metaphor for describing the individual control tasks and their interrelationships. At a maximum, the architecture can greatly simplify the design effort. The simplification is achieved if the different control tasks are 'decoupled' and each task can be carried out using the appropriate reference model. Thus it seems that the value of any architecture will depend upon how well the reference models are constructed.

Judging from the development of the 'open systems architecture' of communication data networks, the design of reference models is of critical importance. The design must evolve from the active participation of all relevant communities. In the case of IVHS this certainly includes transportation agencies, automobile manufacturers, control and communications equipment developers. We hope that forums for such participation that are being established will serve this function.

References

- [1] A. Hitchcock, "Specification of an intelligent vehicle/highway system for investigation of hazard analysis methods," tech. rep., PATH Preprint, University of California, Berkeley, CA 94720, 1989.
- [2] A. Hsu, F. Eskafi, S. Sachs, and P. Varaiya, "Implementing car platoon maneuvers. rev. 2.0," tech. rep., Department of Electrical Engineering and Computer Sciences, Preprint, University of California, Berkeley, CA 94720, November 1990.
- [3] S. Shladover, "Operation of automated guideway transit vehicles in dynamically reconfigured trains and platoons (vol ii)," tech. rep., UMTA Report UMTA-MA-06-0085-79-3, NTIS, Springfield, VA 22161, October 1979.

³⁰We do not suggest inclusion of lane-changing maneuvers because the physical design of a network that permits automatic lane-changing is likely to be significantly different from one in which a single lane is dedicated to fully automated IVHS.

- 4
- CE
- [4] U. Karaaslan, P. Varaiya, and J. Walrand, "Two proposals to improve traffic flow," tech. rep., Department of Electrical Engineering and Computer Sciences, Preprint, University of California, Berkeley, CA 94720, 1990.
 - [5] S. Sheikholeslam and C. A. Desoer, "Longitudinal control of a platoon of vehicles," tech. rep., UCB-ITS-PRR-89-3,6, Institute of Transportation Studies, University of California, Berkeley, CA 94720, 1989.
 - [6] S. A. Smulders, "Filtering of freeway traffic flow," tech. rep., Center for Mathematics and Computer Science, Report OS-R8806, Amsterdam, 1988.
 - [7] S. A. Smulders, "Control of freeway traffic flow," tech. rep., Center for Mathematics and Computer Science, Report OS-R8806, Amsterdam, 1988.
 - [8] A. S. Tannenbaum, *Computer Networks*, vol. 2nd edition. Prentice-Hall, 1988.
 - [9] S. Aggarwal, D. Barbara, and K. Z. Meth, "A software environment for the specification and analysis of problems of coordination and concurrency," *IEEE Transactions on Software Engineering*, vol. 14, pp. 280-290, March 1988.
 - [10] "LANELOCK." in Search. Technical Information Department, GM Research Labs, Warren, MI 48090-9055.
 - [11] B.-W. Zhang, R. Parsons, and T. West, "An intelligent roadway reference systems for lateral vehicle guidance/control," tech. rep., Preprint, Institute of Transportation Studies, University of California, Berkeley, CA 94720, 1989.
 - [12] A. Gollu and P. Varaiya, "Hybrid dynamical systems," in *Proceedings of the 28th Conference on Decision and Control*, (Tampa, FL), pp. 2708-2712, December 1989.
 - [13] J. P. Forestier and P. Varaiya, "Multilayer control of large Markov chains," *IEEE Transactions on Automatic Control*, vol. AC-23, pp. 298-305, February 1978.
 - [14] H. J. Payne, "A critical review of a macroscopic freeway model," in *Engineering Foundation Conference on Research Directions in Computer Control of Urban Traffic Systems*, pp. 251-265, 1979.
 - [15] M. Papageorgiou, *Applications of Automatic Control Concepts to Traffic Flow Modeling and Control*, vol. Lecture Notes in Control and Information Sciences, 50. Springer, 1983.
 - [16] M. Cremer and A. D. May, "An extended model for freeway control," tech. rep., UCB-ITS-RR-85-7, Institute of Transportation Studies, University of California, Berkeley, CA 94720, 1985.
 - [17] M. Tsuzawa and H. Okamoto, "Advanced Mobile Traffic Information and Communication System - AMTICS." Vehicle Navigation and Information System Conference, Toronto, Canada, September 11-13 1989.
 - [18] J. Sparmann, "LISB route guidance and information system: First results of the field trial." Vehicle Navigation and Information System Conference, Toronto, Canada, September 11-13 1989.

- [19] I. Catling and P. Belcher, "Autoguide - Route guidance in the United Kingdom." Vehicle Navigation and Information System Conference, Toronto, Canada, September 11-13 1989.
- [20] H. Al-Deek, M. Martello, A. May, and W. Sanders, "Potential benefits of in-vehicle information systems in a real freeway corridor under recurring and incident induced congestion," tech. rep., UCB-ITS-PRR-88-2, Institute of Transportation Studies, University of California, Berkeley, CA 94720, 1988.
- [21] F. Kobayashi, "Feasibility study of route guidance," *Transportation Research Record*, vol. 737, pp. 107-112, 1979.
- [22] H. Al-Deek and A. May, "Potential benefits of in-vehicle information systems (ivis): Demand and incident sensitivity analysis," tech. rep., UCB-ITS-PRR-89-1, Institute of Transportation Studies, University of California, Berkeley, CA 94720, 1989.
- [23] H. Al-Deek and A. Kanafani, "Some theoretical aspects of the benefits of en-route vehicle guidance," tech. rep., UCB-ITS-PRR-89-2, Institute of Transportation Studies, University of California, Berkeley, CA 94720, 1989.
- [24] L. Bic and A. C. Shaw, *The Logical Design of Operating Systems*, vol. 2nd edition. Prentice-Hall, 1988.