Proceedings of the 2007 IEEE International Conference on Networking, Sensing and Control, London, UK, 15-17 April 2007

Adaptive Dynamic Control for Road Traffic Signals

BG Heydecker, Chen Cai, CK Wong

Centre for Transport Studies, University College London
Gower Street, London WC1E 6BT, England

***Abstract*— Control of road traffic by signals requires timings for the stages during which signal indications remain constant. Approaches to calculation of these durations include traffic-responsive strategies as well as fixed-time ones. In all cases, these control methods require some parameters that will ultimately depend on traffic flows. The present paper introduces an adaptive method for responsive control that adjusts the stage durations according to traffic flows. The methodology used for this is approximate dynamic programming, which uses a state-dependent estimate of future optimised costs to assess decisions. Results are presented for example applications, showing favourable performance by comparison with other approaches.**

I. INTRODUCTION

The control of road traffic at traffic signals provides an effective means to manage conflicting demands for use of road space. In planning and applying control of this kind, decisions are required on the sequence and durations of the periods for which the different signals display green. In order to achieve effective control that provides adequate capacity, and acceptable queue lengths and delays, these control variables should be matched to the flows of traffic. Various strategies have been established for this, ranging from fixed time control in which the sequence and duration of the green indications is the same from cycle to cycle, through heuristic decision approaches based directly upon outputs from traffic detectors, to optimisation approaches based upon dynamic estimates of traffic sates and arrival rates. In the present paper, we report on development of a novel adaptive optimisation-based control strategy that responds to measured variations in traffic flows. This approach is based upon the methodology of adaptive dynamic programming. We present the background to the formulation, the approach that is adopted, and some preliminary results from dynamic microscopic simulation.

II. SIGNAL CONTROL OF ROAD TRAFFIC

Traffic signals work by providing an opportunity for mutually comptible traffic movements to have exclusive use of space in a road junction in turn for periods of time, known as *stages*, that are separated by periods of time during which the individual signals change between green and red, and clearance times elapse between movements losing right of way and those gaining it. While a stage is running, and after a certain *minimum green* time has elapsed for it, the control system can change to a different stage. The structure of the decisions to be made for this control can be specified at each instant by two pieces of information:

- whether or not a change of stage should be initiated;
- if so, which stage it is that should be called next.

If a change of stage is initiated, then this will comit the control actions for a period of several seconds – typically about 12 in the UK, comprising 3 seconds stopping amber, 2 seconds starting amber and 7 seconds minimum green (Ref.[1]). On the other hand, the decision not to initiate a change of stage can be reviewed shortly (*eg* after 0.5 s). These decisions are informed by detector data that relate typically to arrivals that will take place in the next 10 seconds or so.

At simple junctions, the physical layout and the logic of control can limit the number of stages to 2, so that the decision is reduced to whether or not to initiate a call of the other stage. However, at more complicated junctions, or ones where special provision is made of opposed turning traffic (right turners in the UK), then several different stages can be used from time to time according to demand.

The traffic that is controlled by the signals at a single junction will, in general, vary on three distinct timescales. The first of these is at flows vary stochastically from moment to moment due to fluctuations in demand and driver behaviour, flows vary systematically within each day due to peak periods, and flows change over protracted periods of time due to developing demand for travel and traffic management in the vicinity. In practice, each of these separate reasons will apply to some degree.

Fixed-time traffic control systems have been developed (Ref.[2] and [3]) to provide a good allocation of green times. Traffic responsive systems (Ref.[4]) take advantage of stochastic fluctuations in traffic arrivals to allocate green time effectively according to the current and short-term future state. However, both of these approaches require knowledge of the mean arrival rates of traffic in order to calculate plans. In the latter case of responsive control, the way in which the control responds to variations in traffic arrivals can be adapted according to the traffic conditions.

III. DYNAMIC CONTROL

In the present paper, we develop a dynamic approach to control of road traffic based upon Bellman's principal (Ref.[5]) of optimality. We state this formally as:

$$C(\mathbf{x}) = \underset{\psi}{\mathrm{Min}}\left[\underset{\boldsymbol{\omega}}{\mathrm{E}}\left\{c\left(\psi, \mathbf{x}, \boldsymbol{\omega}\right) + e^{-\gamma\tau(\psi)} C\left(\mathbf{y}\left(\psi, \mathbf{x}, \boldsymbol{\omega}\right)\right)\right\}\right] \tag{1}$$

where

$C(\mathbf{x})$ is the total future discounted cost starting from state $\mathbf{x}$,

ψ is the decision (*eg* initiate change of stage or defer decision),

1-4244-1076-2/07/$25.00 ©2007 IEEE

$\mathbf{x}$ is the state of the traffic and the controller,
$\boldsymbol{\omega}$ is information about traffic arrivals,
γ is the discount rate,
$\tau(\psi)$ is the time over which decision ψ is implemented,
$\mathbf{y}(\psi, \mathbf{x}, \boldsymbol{\omega})$ is the state at time $\tau(\psi)$ after decision ψ is implemented when the state is $\mathbf{x}$, and
$c(\psi, \mathbf{x}, \boldsymbol{\omega})$ is the cost of implementing decision ψ.

Several approaches have been developed for this in the past. Ref.[6] used backward dynamic programming formulations in which they used hypothetical knowledge of individual arrivals over a finite time horizon: this approach calculates the future discounted cost $C(\mathbf{x})$ recursively backwards from a future terminal time, using the values that have already been calculated in the right-hand side of Bellman's equation. Although they recognised that this implies an impractical data requirement, they promoted this approach as providing an absolute minimum of cost for the arrival pattern that arises; they showed that the optimal decisions in the short run were insensitive to variations in traffic arrivals at times after about 25 seconds into the future. Ref.[7] developed the OPAC rolling optimisation procedure that uses a direct search method over estimates of delay based upon detected arrivals for the short-term future and estimated arrivals thereafter using an expression for $C(\mathbf{x})$ that is linear in queue lengths. The value of an expression of this kind will depend on the mean rate of arrivals in the future.

In the present approach, we consider future discounted delays over an infinite planning horizon. However, rather than estimate these backwards in time, they are estimated endogenously using a method of approximate dynamic programming (ADP), as in [8]. In this case, approximate values of the cost function $C(\mathbf{x})$ are used on the right-hand side of Bellman's equation to inform decisions. The resulting optimised value of $C(\mathbf{x})$ on the left-hand side is then used to update the approximation for use in the next decision. This approach requires either a discrete state space for which the values can be tabulated and updated, or a parametric form for the function $C(\mathbf{x})$. In the present approach, we adopt a linear form

$$C(\mathbf{x}) = \alpha \sum_{i \in G} x_i + \beta \sum_{j \in R} x_j \tag{2}$$

where R is the set of streams that have a red indication, G is the set of streams that have a green indication, x_i is the queue length in stream i, and $\mathbf{A} = (\alpha, \beta)$ is a vector of parameters. The information that is used to estimate the parameters $\mathbf{A}$ enters in the term $c(\psi, \mathbf{x}, \boldsymbol{\omega})$ and thus depends on the arrival rate that is manifested in $\boldsymbol{\omega}$. Through this mechanism, the way in which the control responds to traffic will be adapted according to the prevailing flows as the parameters $\mathbf{A}$ are adjusted. This form of ADP therefore corresponds to an adaptive learning approach.

IV. The Adaptive Control Algorithm

A. A two-stage case

1) Definition of the state variables

The two-stage problem is the simplest case for an isolated road junction. The two incompatible stages are either on green or red. A 5-second-time increment is adopted throughout the examples presented here, with interstage time and minimum green time each equal to on time increment. The simplicity of this formulation allows the state variable $\mathbf{x}$ to be expressed as:

$$\mathbf{x} = \left[\mathbf{L}^T, \mathbf{G}^T\right]^T .$$

Thus the state variable $\mathbf{x}$ can be seen as the concatenation of two subordinate state variables — the traffic state variable $\mathbf{L}$, and the controller state variable $\mathbf{G}$, where

$$\mathbf{L} = \begin{bmatrix} L_1 \\ \vdots \\ L_I \end{bmatrix}, \text{ where } L_i \text{ is the queue length on stream } i;$$

$$\mathbf{G} = \begin{bmatrix} G_1 \\ \vdots \\ G_I \end{bmatrix}, \text{ where } G_i = \begin{cases} 1 \text{ if stream receives green,} \\ 0 \text{ if stream receives red.} \end{cases}$$

2) The state transfer function y

Function $\mathbf{y}(\psi, \mathbf{x}, \boldsymbol{\omega})$ transfers the current state to the next state after time $\tau(\psi)$. Additional to primary state variable $\mathbf{x}$, it depends on traffic arrivals $\boldsymbol{\omega}$, decision ψ, and the state transfer mechanism in the system.

Traffic arrival information vector $\boldsymbol{\omega}$ and decision vector ψ are introduced as:

$$\boldsymbol{\omega} = \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_I \end{bmatrix},$$

and

$\psi = 1$ for an immediate change of signal,
$\psi = 0$ to defer the decision on changing.

Given time t, state $\mathbf{x}_t$, arrival traffic information $\boldsymbol{\omega}_t$, and decision ψ_t, the transfer function for subordinate state variable $\mathbf{L}$ is defined as:

$$\mathbf{L}_{t+\tau(\psi)} = \mathbf{L}_t + \boldsymbol{\omega}_t - \mathbf{O}_t\left(\psi_t, \mathbf{x}_t, \boldsymbol{\omega}_t\right),$$

where the $\mathbf{O}$ is the outflow vector, given by

$O_i = G_i\, Min(s_i,\ L_i + \omega_i)$ and s_i is the saturation departure rate for stream i.

The control part of the state variable $\mathbf{G}$ is transformed by:

$$\mathbf{G}_{t+\tau(\psi)} = (\mathbf{G}_t + \psi_t)_{\mathrm{mod}\,2}.$$

The primary state variable $\mathbf{x}$ is then calculated according to associated with its two subordinate variables $\mathbf{G}$ and $\mathbf{L}$. The function $\mathbf{y}(\psi, \mathbf{x}, \boldsymbol{\omega})$ can be summarized as:

$$\mathbf{y}(\psi_t, \mathbf{x}_t, \omega_t) = \begin{cases} L_{t+\tau(\psi)} = L_t + \boldsymbol{\omega}_t - O_t(\psi_t, \mathbf{x}_t, \boldsymbol{\omega}_t), \\ G_{t+\tau(\psi)} = (G_t + \psi_t)_{\mathrm{mod}\,2} \end{cases} \tag{3}$$

3) The adaptive value function approximation

Equation (2) and (3) together enable the calculation of $C(\mathbf{y}(\psi, \mathbf{x}, \boldsymbol{\omega}))$ on the right-hand side of (1). Also on the right-hand of (1), the cost of implementing decision ψ is approximated by:

$$c(\psi, \mathbf{x}, \boldsymbol{\omega}) = \sum_{i=1}^{I} \left[L_i + \left(\boldsymbol{\omega}_i - O_i(\psi_t, \mathbf{x}_t, \boldsymbol{\omega}_t) \right) / 2 \right] \tau(\psi). \tag{4}$$

The total future discounted cost $C(\mathbf{x})$ can then be estimated from (1), using (2) - (4). The adaptive learning feature of the ADP strategy requires obtaining $C(\mathbf{x})$ and then estimating the new observation of parameters $\mathbf{A}$ in the linear approximation (2) of the future discounted cost by:

$$\hat{\alpha}^k = \frac{\partial C}{\partial L_g} = \frac{C(\mathbf{x}_t + \Delta \mathbf{x}_g) - C(\mathbf{x}_t)}{\|\Delta \mathbf{x}_g\|}, \tag{5}$$

$$\hat{\beta}^k = \frac{\partial C}{\partial L_r} = \frac{C(\mathbf{x}_t + \Delta \mathbf{x}_r) - C(\mathbf{x}_t)}{\|\Delta \mathbf{x}_r\|}, \tag{6}$$

where $\Delta\mathbf{x}$ is a column vector with Δx in the respective entry and the zero elsewhere. The new observations are used to update current parameters through:

$$\alpha^k = (1 - \theta_k)\alpha^{k-1} + \theta_k \hat{\alpha}^k, \tag{7}$$

$$\beta^k = (1 - \theta_k)\beta^{k-1} + \theta_k \hat{\beta}^k, \tag{8}$$

where θ is the stepsize and k the counter of updates. In the present work, we have used the MSA rule $\theta_k = 1/k$.

4) The ADP algorithm

The ADP algorithm can now be shown as in Fig.1.

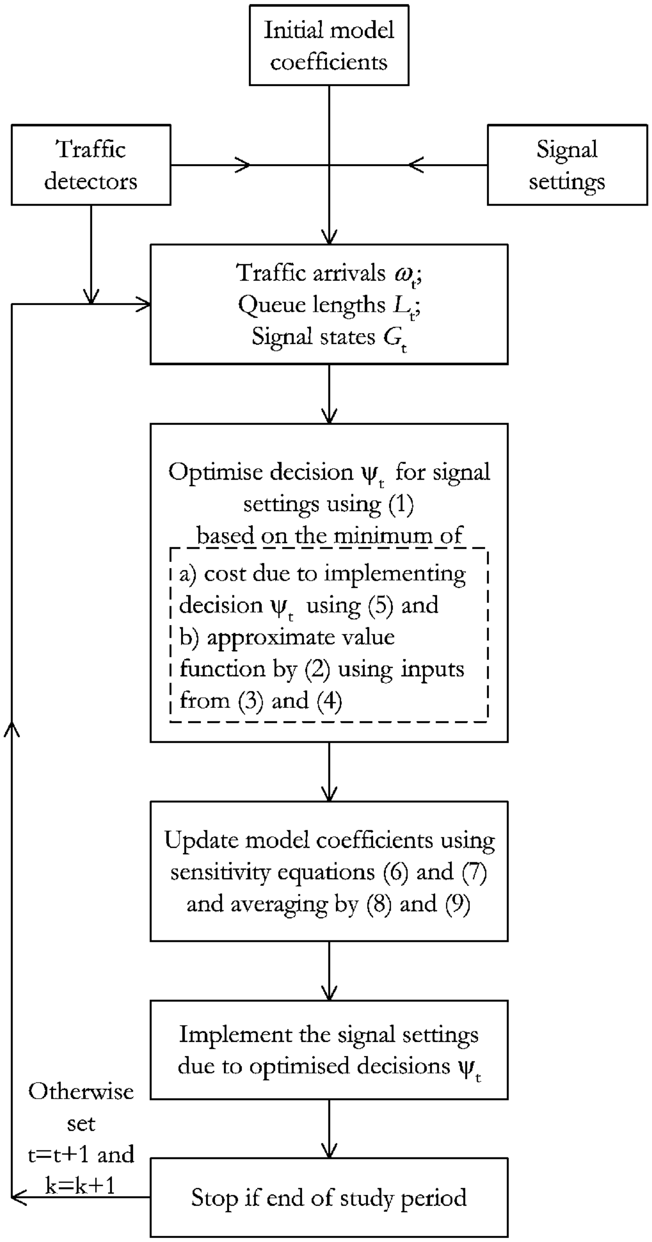

Fig.1. The ADP algorithm

B. A multi-stage case

To extend the ADP strategy to a multi-stage junction, we extend the state description accordingly and allow stream-specific weights α and β in the approximate cost function (2). The state variable $\mathbf{x}$ is augmented to include separate entries for each stream. Given a total of I mutually incompatible traffic streams, and thus I control stages, for each stream i we introduce the traffic state variable L_i as in the two-stage problem and a corresponding controller state variable G_i, where

$$G_i = \begin{cases} 1, \text{ if stream } i \text{ receives green}, \\ 0, \text{ if stream } i \text{ receives red}. \end{cases}$$

In the approximate value function, each stream will have two parameters to represent whether it receives a red or a green indication. The two parameters are elements in vector A_i, where

$$\mathbf{A}_i = \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}.$$

As in the two-stage problem, α_i is a coefficient in the cost function of the queue length on stream i if it receives green, whilst β_i is the corresponding coefficient for that stream on red. An advantage of introducing the stream-specific coefficients $\mathbf{A}_i$ instead of using generic ones is that the particular properties of each stream can be represented indirectly through them. The approximate value function is then given by

$$C(\mathbf{x}) = \sum_{i=1}^{I} \left[G_i \alpha_i + (1 - G_i) \beta_i \right] L_i \,. \tag{9}$$

Equation (9) will be used in (1) and the decision rules will be further expanded to accommodate the additional options as to which stage to call. Value function parameters will be updated in the same way as in the two-stage problem.

If an ADP strategy is able to work at a multi-stage junction, it then can be extended to coordinate signal control by adopting the concept of decentralized network control as in [7] and [9]. This means the network controller finds a critical junction in the network at each epoch to synchronize critical variables like minimum and maximum stage length, whereas performances are optimized by local controller at individual junctions with the subjection to the limits set by network controller. Sophisticated traffic models are required for the network dynamic traffic signal coordination, and there are many choices for them.

V. EXAMPLE CALCULATIONS

A. *Experiment on Two-Stage Control*

In order to illustrate use of this approach, we have undertaken calculations for a simple example junction. In this case, there are 2 approaches and hence 2 signal stages. We undertook 10 simulations with each of 8 combinations of constant mean arrival rates and evaluated 4 different control strategies:

Backward dynamic programming (Robertson and Bretherton) (BDP),
Approximate dynamic programming with linear value function approximation (ADP),
Robertson and Bretherton's empirical near-optimum strategy (RB), and
Gartner's optimum sequential constrained algorithm (OPAC).

The results of this are shown in the Table 1 below. The BDP method, here implemented after Robertson and Bretherton's DYPIC, provides the optimal (if unrealisable) performance for each sequence of arrivals that is simulated, and therefore provides a lower bound for all delays. The performance of the ADP algorithm is in many cases better than that of the other approximate methods investigated here.

TABLE 1: Simulation results: Mean total rate of delay over 10 minutes (mean of 10 runs; vehicle-seconds per second)

	Flow (veh/h)	Arm A			
Method	Arm B	252	396	600	678
BDP	240	0.51	0.88	1.52	1.86
ADP		0.59	1.03	1.98	2.39
RB		0.65	1.14	1.90	2.31
OPAC		0.69	1.13	2.04	2.53
BDP	432	0.99	1.73	3.36	4.38
ADP		1.22	2.13	3.76	4.95
RB		1.21	2.10	3.78	4.91
OPAC		1.32	2.19	4.05	5.34

The adaptive estimates of the parameters α of the total cost function are shown in Fig.2. These show good convergence within the time frame of the simulation, with ultimate values of about $\alpha = (1.92, 2.85)$ for a single case with of mean arrival rates (678, 432) vehicles/h. These results show that the control method proposed here has the capability to adapt the parameter values according to prevailing traffic flows. The resulting control performance of 4.95 vehicle-seconds per second is good by comparison with the other realistic traffic responsive control methods, which achieve 4.91 and 5.34 veh-s/s

.

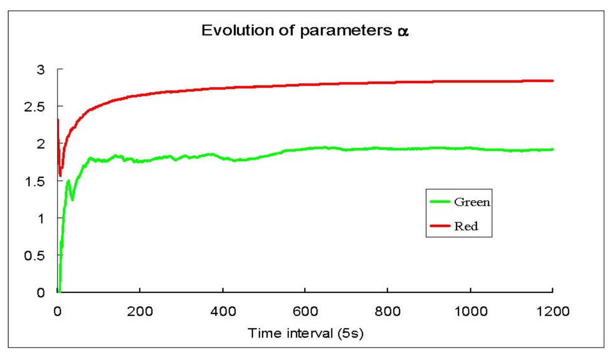

Fig.2: Evolution of parameters of cost function approximation

B. *Experiment on Multi-Stage Control*

Road junctions are often controlled using more than two stages, and to investigate the capability of the ADP this more realistic multi-stage control environment, an experiment was extended to represent a three-stage case. The comparable control strategy in the three-stage case is the best fixed-time control.

Each of the three stages serves a single traffic stream, has a minimum green time of 5 seconds, and is followed by an interstage time that has effective duration 5 seconds. Each pair of the three streams is mutually incompatible. The mean arrival rates in the streams are:

Stream 1: 432 v/h;
Stream 2: 252 v/h;

Stream 3: 432 v/h,

The total demand of 1116 vehicles/h in the three streams represents a near saturation level of traffic with respect to the common saturation flow of 1440 vehicles/h: in this case the flow ratio is $Y = 1116/1440 \approx 0.775$.

The best fixed-time control strategy is found through direct search of stage durations and stage sequences. Once the stage durations and sequence have been determined, they are kept fixed from cycle to cycle. This is unlike the ADP control in which the stage length and sequence may vary in response to prevailing traffic flows.

The delay performance, averaged over 10 runs, each of duration 10 minutes at a three-stage junction, of ADP is 10.01 vehicle-seconds per second. The corresponding performance of the best fixed-time control is 19.17 vehicle-seconds per second, achieved with effective stage durations and sequence given below:

Stage 1: Stream 2, 25 seconds;
Stage 2: Stream 3, 40 seconds;
Stage 3: Stream 1, 40 seconds.

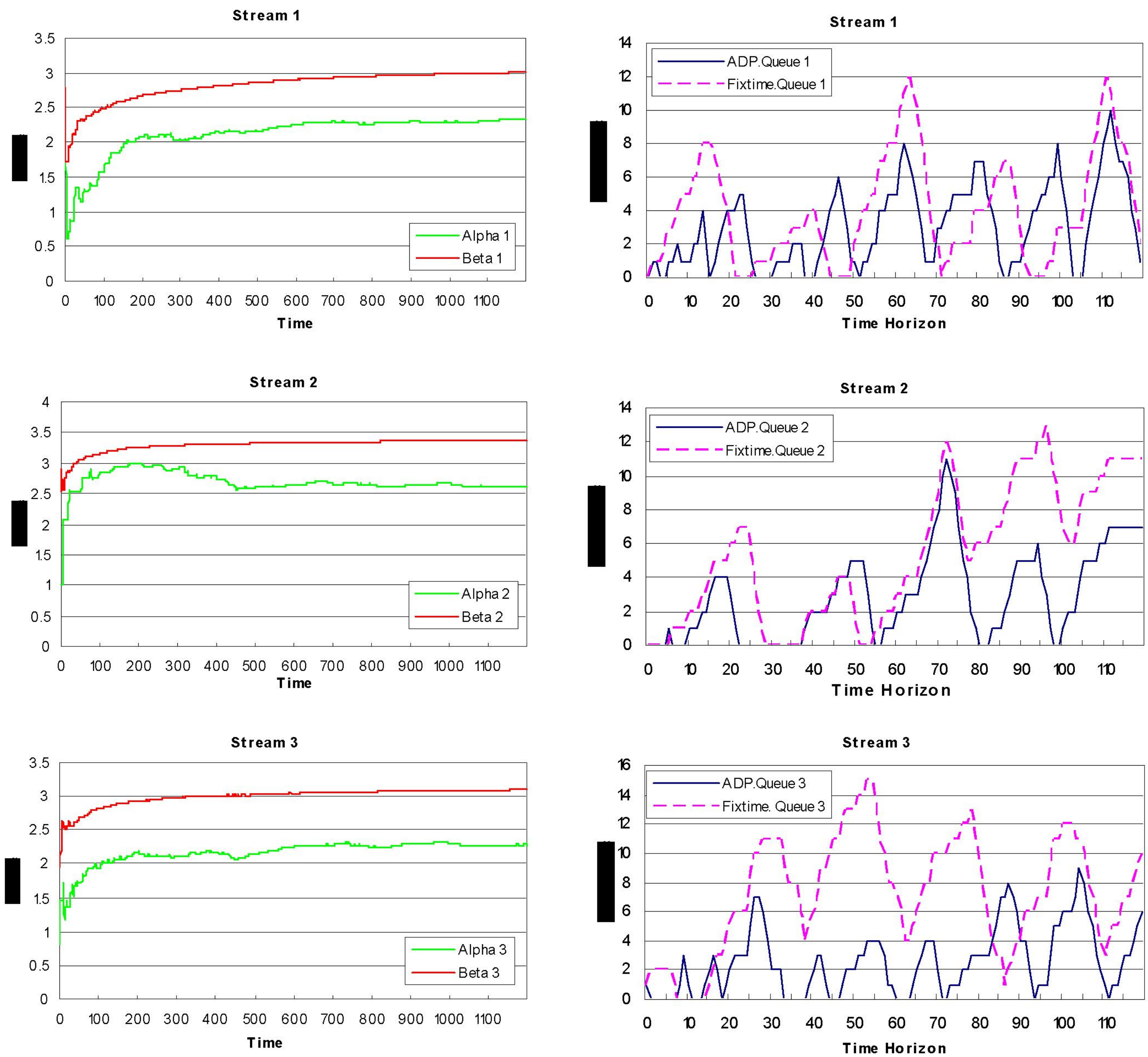

Fig.3. The convergence of function parameters and performance comparison between ADP and Fixed-time control in the first 10 minutes

Adding the effective intergreen to the stage durations gives a cycle time of 120s, which corresponds to the maximum that was allowed.

There is an approximately 48% savings in delay from ADP over the best fixed-time control, and the savings are earned by responding to short-term variations in arrivals. The ADP strategy, compared with fixed-time control, does not have a fixed sequence or cycle time. The stages that make up the cycle in a fixed sequence in fixed-time control, are called as and when required and with durations determined by traffic conditions. These features allow the traffic signal controller to change signal indication more swiftly among traffic stages when short-time traffic is light and queues in the system are rare. On the other hand, when traffic is heave and queues are prominent, controller will extend the current stage in green to dissipate the queues as much as possible, *ie* to maximize the capacity of the system. In the meantime, the adaptive approximation of value function in ADP strategy keeps evaluating the function parameters so that it is able to reflect the delay incurred in the further by adopting current decision more appropriately.

The adaptive estimation of function parameters in multi-stage control problem are shown in Fig.3. There is good convergence, and the exact values to which parameters converge are tabulated in Table 2. The number of times that each parameter is updated during the 10 minutes of simulated time is also noted. Fig.2. also illustrates the differences in control and performance between ADP and fixed-time strategy in terms of the queues on each stream over time. As has just been discussed, the ADP controller allocates more green time to streams 1 and 3, which have the higher mean arrival rate, than fixed-time controller does, and ADP controller leaves little residual queue at the end of the stage in green.

TABLE 2 Converged values of function parameters in multi-stage control

	Initial		Final		Updates	
Stream	α_i	β_i	α_i	β_i	α_i	β_i
1	1	2.5	2.33	3.01	454	746
2	1	2.5	2.63	3.37	277	923
3	1	2.5	2.28	3.10	469	731

VI. Conclusion

The methods presented in this paper show the potential for approximate dynamic programming in on-line control of road traffic signals. The central feature of this approach is that it uses an approximate expression for the cost of future delays associated with a controller and traffic state following a decision interval. This approach obviates the need for explicit calculations of optimal minimized future discounted costs, which is appropriate in the absence of even medium-term arrivals data. Furthermore, the ADP approach estimates a suitable cost function for this use on-line.

The proposed method can be extended readily to junctions that are controlled by several stages, thus overcoming the limitations associated with earlier advanced traffic responsive control formulations. In the present case, example calculations for a junction with 3-stage control, 5s discrete time control decisions, and fixed mean arrival rates that load the junction fairly heavily, the proposed method leads to delays that are about half of those that are achieved by fixed-time control. This establishes the flexibility and benefits of the ADP approach.

The present approach uses a pre-determined sequence of weights in the updating process for the parameters of the cost function, which is appropriate for fixed mean arrival rates in the streams. This approach could be extended readily to weights that are modified so as to adapt to changes in mean arrival rates so that the parameter values of the cost function can be updated more rapidly when appropriate.

Acknowledgements

The second author would like to thank the Rees Jeffreys Road Fund for financial support for the work presented here. The third author would like to thank The Croucher Foundation for their support of the present work during the Croucher Foundation Fellowships 2006-8.

References

[1] DoT TR 0141: Microprocessor-based traffic signal controller for isolated, linked and urban traffic control installations. London: Department of Transport, 1991.

[2] Allsop, R.E. SIGSET: a computer program for calculating traffic signal settings. *Traffic Engineering and Control*, **13**(2), 58-60, 1971.

[3] Allsop, R.E. SIGCAP: a computer program for assessing the traffic capacity of signal-controlled road junctions. *Traffic Engineering and Control*, **17**(8/9), 338-41, 1976.

[4] DoT MOVA system of traffic contral at signals. DoT Highway Safety and Traffic Departmental Standard, TD 35/91. London: DETR, 1991.

[5] Bellman, R. Dynamic Programming. Princeton University Press, 1957.

[6] Robertson, D.I., and Bretherton, R.D. Optimum control of an intersection for any known sequence of vehicular arrivals. Proceedings of the 2nd IFAC-IFIP-IFORS Symposium on Traffic Control and Transportation Systems, 1974.

[7] Gartner N.H. OPAC: a demand-responsive strategy for traffic signal control. *Transportation Research Record*, **906**, 75-84, 1983.

[8] Cheung, R and Powell, WB. An algorithm for multistage dynamic network with random arc capacities, with an application to dynamic fleet management. *Operations Research*, **44**(6), 951-63, 1996.

[9] J. J. Henry, "PRODYN tests and future experiments on ZELT," VNIS '89: *Vehicle Navigation and Information Systems, IEEE Conference, Toronto,* 1989.