MARKOVIAN DECISION CONTROL FOR TRAFFIC SIGNAL SYSTEMS

Xiao-Hua Yu, Allen R. Stubberud
xhyu@ece.uci.edu, arstubbe@uci.edu
Department of Electrical and Computer Engineering
University of California, Irvine
Irvine, CA 92697

ABSTRACT

Setting signals at intersections to minimize the queue length and vehicle delay time is a key goal in traffic management. In this paper, a new control strategy for a signalized traffic intersection is developed by applying Markovian decision control theory. Statistical analysis of simulation results with different arrival rates indicate the excellent potential of this approach.

I. INTRODUCTION

The traditional approach for traffic signal control is to employ a set of fixed-time signal timing plans which are generated off-line based on the deterministic traffic conditions during different time periods of the day (e.g., peak hours, off-peak hours). For example, the most widely used computer software package for traffic signal control, TRANSYT-7F (TRAFFIC NETWORK STUDY TOOL) [3] relies on historical data and is considered to be an effective off-line control strategy. In SCOOT (SPLIT, CYCLE and OFFSET OPTIMIZATION TECHNIQUE) [4] and SCATS (SYDNEY COORDINATED ADAPTIVE TRAFFIC SYSTEM) [5], the control strategy is to "match" the current traffic conditions obtained from detectors to the "best" precalculated off-line timing plan. Generally, these conventional traffic signal control approaches do not hold much promise to achieve fully real-time adaptive control.

The Markov decision process, or the controlled Markov process, has been studied by many researchers since the 1950s, e.g. [1]. It has found applications in many areas. A discrete time, stationary Markov control model is defined on (X, A, P, R) where X: the state space, where every element $x \in X$ is called a state; A: the set of all possible controls (or alternatives); P: a probability measure space, in which an element $p_{i,j}^{k}$ denotes the transition probability from state i to state j under alternative k; and R: a measurable function, called a one-step reward.

Choosing a particular alternative results in an immediate reward and a transition probability to the next step. The ultimate objective is to find the supremum (least upper bound) of the total expected discounted reward over an infinite period of time:

$$J \triangleq E\left[\sum_{t=0}^{\infty} \beta^{t} r(x_{t}, a_{t})\right]$$

where r is the one-step transition reward, β ($0 \le \beta < 1$) is the discount factor, and a is the policy. The optimal reward v^* is defined as:

$$v^*(x, a^*) = \sup_{a \in A}[J(x,a)]$$

It can be obtained by solving a DPE (dynamic programming equation):

$$v^* = Tv^*,$$

where T is a contraction mapping and:

$$Tv(x) = \max_{a \in A} \left[r(x,a) + \beta \sum_{j=1}^{N} v(x) p_{i,j}^{a} \right].$$

It has been proved that the optimal solution of the above DPE is unique and can be calculated iteratively by the successive approximation method [1]:

$$v_n(x) = \max_{a \in A} \left[r(x,a) + \beta \sum_{j=1}^{N} v_{n-1}(x) p_{i,j}^{a} \right]$$

Therefore, for a specific control problem, once the transition matrix and the reward matrix are determined, then by maximizing the total expected reward, a policy of choosing a certain alternative in each state will be obtained. This represents the optimal strategy which should be taken.

In the following sections, a new approach based on the above Markov decision theory is proposed and applied to traffic signal control problem.

II. TRAFFIC DYNAMIC MODEL FOR AN INTERSECTION

Modeling traffic dynamics and optimizing the control signal are two interrelated problems. A typical four-legged intersection is shown in Fig. 1. There are four approaches in this intersection and each one of them has one through movement and one left turn movement, summing to a total of 8 movements. The number on each movement is labeled according to NEMA (National Electrical Manufacturers Association) convention.

Assume a continuous traffic flow process that is sampled every Δt time interval with the discrete time index, k. The output of the intersection (i.e., number of vehicles leaving this intersection) $\underline{q}_{out}(k)$ can be defined as a vector:

$$\underline{q}_{out}(k) = \left[q_{out}^{1}(k), q_{out}^{2}(k), ..., q_{out}^{8}(k) \right]^{T}$$

where the superscript j (j = 1, 2, ..., 8) denotes the j-th movement. Similarly, the current queue $\underline{q}(k)$ will be defined as:

$$\underline{q}(k) = \left[q^{1}(k), q^{2}(k), ..., q^{8}(k) \right]^{T}$$

$\underline{q}_{out}(k)$ can be further expressed as a function of the current control of the intersection, $\underline{u}(k)$, and $\underline{q}(k)$:

$$\underline{q}_{out}(k) = \underline{f}_{out} \left(\underline{u}(k), \underline{q}(k) \right)$$

where $\underline{f}_{out}(k)$ is also a vector:

$$\underline{f}_{out}(k) = \left[f_{out}^{1}(k), f_{out}^{2}(k), ..., f_{out}^{8}(k) \right]^{T}$$

and

$$f_{out}^{j}(\bullet) = \begin{cases} \min\left[q^{j}(k); \dfrac{\Delta t}{h_{min}} \right] & \text{when } u^{j}(k) = 1 \\ \\ 0 & \text{when } u^{j}(k) = 0 \end{cases}$$

where j = 1, 2, ..., 8. h_{min} is the minimum headway, and $u^{j}(k)$ is the control signal for the jth movement: $u^{j}(k) = 1$ denotes a green signal, $u^{j}(k) = 0$ indicates a red signal.
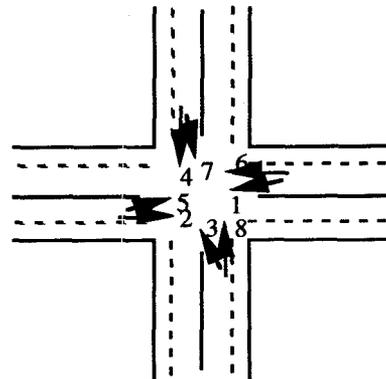
Figure 1. A typical traffic intersection

The current queue $\underline{q}(k)$ can also be written as:

$$\underline{q}(k) = \underline{q}(k-1) + \underline{q}_{in}(k) - \underline{q}_{out}(k)$$

where $q(k-1)$ is the queue at the previous time instant $(k-1)$ and $q_{in}(k)$ is the input (number of vehicles) during time interval $[k-1, k)$.

The time duration of the current signal, τ, must be bounded between some minimum and maximum time period:

$$\tau_{min} \leq \tau \leq \tau_{max}$$

In an eight-phase dual-ring control, the phases are divided into two groups (rings) by a barrier. In each ring, 4 movements (2 through movements and their corresponding left turn movements) must be served if there is demand. Theoretically, there are $2 \cdot 4! = 48$ different phase sequences available, but in fact, in order to avoid conflict traffic, only certain sequences (10 out of 48) are allowed (see reference [6] for more details). Since there are up to 3 admitted phases in each ring, when choosing current control, three previous control signals need to be considered in order to satisfy the sequence constraint:

$$\underline{u}(k) = \underline{f}_u \big(\underline{q}(k), \tau, \underline{u}(k - \tau_1), \underline{u}(k - \tau_2), \underline{u}(k - \tau_3) \big)$$

III. TRAFFIC SIGNAL CONTROL USING MARKOV DECISIONS

To apply Markovian control to traffic systems, a state space X and a probability measure P must be defined. A threshold (number of vehicles) is chosen for the queue of each movement at an intersection. If the queue length of a specific movement is greater than the threshold value, then this movement is defined to be in its congestion mode; otherwise it is in the non-congestion mode. These two modes (congestion/non-congestion) are defined as the two states in the state space X. The signal phasing can be considered as different alternatives in each state. For a simple example, assume the traffic flow moves only in two directions: either north/south (denoted by 1) or

east/west (denoted by 2). Then there are four possible situations: a) both directions are noncongested; b) direction 1 is congested but direction 2 is noncongested; c) direction 2 is congested but direction 1 is noncongested; and d) both directions are congested. These four different situations can be defined as the four states of the Markov process. Furthermore, if there are 8 independent movements under 8-phase signal control, the traffic control problem can be formatted as a 256-state Markov process with 8 alternatives in each state.

The state space is discrete, thus the probability measure P is defined as a discrete transition law. An element of this matrix P, i.e., $p_{i,j}^k$ denotes the transition probability from state i to state j under alternative k. Let $\pi(k)$ be a row vector of state probabilities (i.e., $\pi_i(k)$ is the probability that the system will occupy state i after k transitions). In the traffic control problem, the probability matrix P is time-varying due to the time-varying traffic flow, therefore:

$$\underline{\pi}(k + 1) = \underline{f}_\pi [\underline{\pi}(k), \underline{P}(k)]$$

where the probability matrix $\underline{P}(k)$ is a function of the current queue, the estimated number of arrivals in the next time interval, and the control signal:

$$\underline{P}(k) = \underline{f}_p [\underline{q}(k), \hat{\underline{q}}_{in}(k + 1), \underline{u}(k), q_g]$$

The probability matrix can be specified based on different arrival patterns. Under most circumstances, the arrival of vehicles at an isolated intersection follows the Poisson distribution, i.e.,:

$$p(n) = \frac{(\lambda \Delta t)^n e^{-\lambda \Delta t}}{n!}$$

where $n = 1, 2, ...$, λ is the arrival rate and Δt is the time interval. Assuming that at a specific time instant, the current queue length of a specific movement i is denoted by q, and there are q_g vehicles passing through the intersection if the signal of this direction is green, then:

$$p_{S_i \to N_i}^{u_i} = p(\hat{q}_{in}^i + q^i - \delta(u_i)q_g^i \le q_{threshold}^i)$$

and

$$p_{S_i \to C_i}^{u_i} = 1 - p_{S_i \to N_i}^{u_i}$$

where

$$\delta(u_i) = \begin{cases} 1, & \text{when } u_i = G_i \\ 0, & \text{otherwise} \end{cases}$$

and $S_i = N_i, C_i$ (N_i for non-congestion and C_i for congestion); $u_i = G_i, R_i$ (G_i for green signal and R_i for red signal).

The reward matrix R has the same dimension and a similar definition to the probability matrix. The control objective herein is to minimize the queue length, so the functions of queue length corresponding to different states are chosen to generate the reward matrix:

$$R_{state1, state2}^{u_i} = f_u(q_0^i, q_{threshold}^i, u_i)$$

Once the transition matrix and the reward matrix are obtained, a certain policy of choosing a certain alternative in each state, which is the optimal strategy we should take, will be obtained by maximizing the total expected reward. It has been shown that this optimal solution is unique and can be calculated iteratively [1]. Thus, the problem of choosing signal phasing becomes a decision-making problem for a Markov process.

The adaptive control procedure for a traffic intersection consists of 2 parts, i.e., a probability calculation and a Markov decision based on the probability and reward (Fig. 2).

Being related to the current state of each traffic movement, the probability matrix and reward matrix are time-varying in the traffic control problem. Future arrival information is needed for adaptive control; however, it is very difficult to make a long term estimate due to the randomness of traffic system. Thus the sampling frequency of the traffic system should be set as high as possible. On the other hand, a large sampling rate will increase the cost and

computation time. Here we choose $\Delta t = \tau_{mini}$ (i.e., minimum green extension time). Every Δt seconds, the time-varying probability matrix P and reward matrix are calculated; then a decision is made to choose the control signal for the next time interval based on the current measurement from the detector and our estimation. Once the optimal policy is found, it is only implemented for Δt seconds. At the next time interval, the probability matrix and reward matrix are updated and the whole decision-making process is repeated.
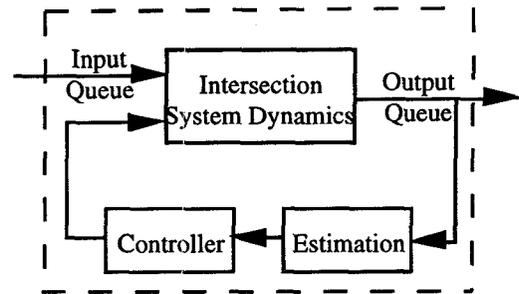
Figure 2. Block diagram of traffic control system for one intersection

IV. SIMULATION RESULTS

The proposed Markovian adaptive control algorithm was simulated for an isolated intersection (with a Poisson arrival pattern generated as the external input) to evaluate its performance in comparison to a fully actuated control method. Some of the simulation parameters are summarized as the following:
Minimum green time: 3 seconds
Maximum green time: 30 seconds
Extension (gap) time: 3 seconds
Yellow time: 3 seconds
Minimum departure headway: 2 seconds

The two algorithms were tested on four different arrival rates, i.e., 200 vehicles per hour per movement, 300 vehicles per hour per movement, 400 vehicles per hour per movement, 500 vehicles per hour per movement, and 600 vehicles per hour per movement. For each arrival rate, the algorithms were tested on forty different sets of random data. The mean, covariance and standard deviation of the average steady state delay (of the 40 sets of data) were calculated and are listed in table 1, where "MAC" stands for the Markov adaptive control algorithm, and "FAC" stands for the fully actuated control. The means (of the 40 sets of data) of the steady state delay are also plotted in Fig. 3, where the solid line represents the Markov algorithm and the dotted line represents the fully actuated control.

	200		300		400		500		600	
	FAC	MAC	FAC	MAC	FAC	MAC	FAC	MAC	FAC	MAC
Mean	11.25	11.51	16.61	12.26	29.60	13.34	41.64	18.27	68.09	53.73
Cov.	2.50	4.12	6.94	3.10	22.63	4.10	13.22	8.62	92.44	226.15
Std.	1.58	2.23	2.64	1.76	4.76	2.03	3.64	2.94	9.61	15.04

Table 1. Mean, covariance, and standard deviation of two algorithms

	200		300		400		500		600	
	FAC	MAC	FAC	MAC	FAC	MAC	FAC	MAC	FAC	MAC
Lower limit	8.6	7.22	12.2	9.46	21.7	9.56	34.4	13.15	51.5	19.62
Upper limit	14.6	15.20	24.2	16.40	42.3	18.75	48.0	25.62	91.5	96.67

Table 2. Bounds for simulation results
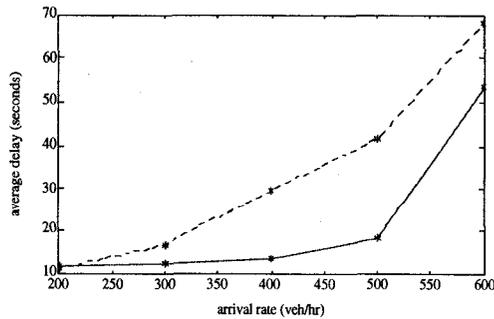
Figure 3. Mean of two algorithms
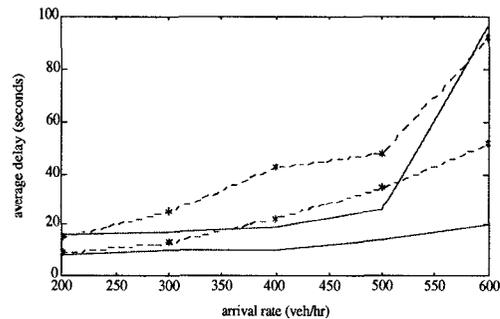
Figure 4. Bounds for simulation results

4786

Using the concept of distribution-free order statistics, the limits within which at least 90% of the probability of the steady state delay obtained from simulation are found in Table 2, with 92% confidence. In other words, the probability that 90% of the probability of the delay time lies between the above lower and upper bound is 0.92. The bounds are also plotted in Figure 4.

From the above data, it is shown that when the traffic volume is slight (e.g., arrival rate is 200 vehicles/hour/movement), the performance of Markov algorithm is comparable with the fully actuated controller. However, when the traffic volume increases, especially for arrival rate (λ) equal to 400 (veh./hr/movement) and 500 (veh./hr/movement), the Markov algorithm outperforms the traditional one. For example, when λ =300, the Markov algorithm shows about 26.19% improvement on the average steady state delay. When λ =400 and λ =500, the average steady state delay of the Markov controller is only about one half of that of the fully actuated controller. When λ is further increased, the intersection becomes saturated, resulting a large delay for both algorithms.

V. CONCLUSIONS

The traffic system is a stochastic system. In this paper, a new approach for traffic signal control based on Markov decision theory is presented. Computer simulation results and analysis are also reported. From the simulation, the new approach is seen to outperform the traditional fully actuated control. Further evaluation and testing are being performed.

REFERENCES

[1] Hernandez-Lerma, O., *Adaptive Markov control process*, Applied mathematical sciences 79, Springer-Verlag, 1989

[2] Recker W., Ramanathan, B., Yu, X.-H., et. al., "Markovian real-time adaptive control of signal systems", *Journal of mathematical and computer modeling*, Pergamon Press, vol. 22, August 1995

[3] Transportation research center, *Traffic network study tool: TRANSYT-7F software summary*, University of Florida, 1987

[4] Hunt, P. B., Robertson, D. I., et al., "The SCOOT on-line traffic signal optimization technique", *Traffic engineering and control*, April 1982

[5] Lowrie, P., "The Sydney coordinated adaptive control system -- Principles, methodology, algorithms", *IEE conference publication*, vol. 207, 1982

[6] Wilshire, R., Black, R., et al., *Traffic control systems handbook*, FHWA-IP-85-12, 1985