Traffic signal optimization via Markovian decision processes

Jay Prakash [1], **Neeti Sharma** [2]

[1] Assistant Professor, Department of BCA (Mathematics), GNIOT Institute of Professional Studies, Greater Noida, Uttar Pradesh, India

[2] Professor, Department of BCA (Mathematics), GNIOT Institute of Professional Studies, Greater Noida, Uttar Pradesh, India

* Corresponding Author: **Jay Prakash**

Article Info

ISSN (online): 2582-7138
Volume: 06
Issue: 03
May-June 2025
Received: 09-03-2025
Accepted: 10-04-2025
Page No: 410-414

Abstract

Traffic signal optimization is a critical component of modern urban traffic management systems. This paper proposes a novel approach to optimizing traffic signal control using Markovian Decision Processes (MDPs). By modeling the traffic flow at intersections as a stochastic process, the Markovian framework allows for the development of decision policies that minimize congestion and improve traffic flow. The system considers various factors such as traffic volume, waiting times, and signal timings, incorporating these into a dynamic model that adapts to real-time traffic conditions. A reinforcement learning algorithm is applied to iteratively improve the control policy, aiming to minimize the total delay and improve throughput across the network of traffic signals. Experimental results demonstrate that the proposed MDP-based approach outperforms traditional traffic signal control strategies, providing significant reductions in average waiting time and overall traffic congestion. This study highlights the potential of advanced decision-making models in creating more efficient, responsive traffic management systems in urban environments.

Keywords: Markovian Decision Processes (MDPs), Traffic signal optimization, stochastic process, Dynamic model

1. Introduction

The efficient management of traffic signals is essential for optimizing the flow of vehicles in urban transportation networks. As cities grow and traffic congestion increases, traditional traffic signal control methods often fail to meet the dynamic needs of modern traffic conditions. To address this challenge, advanced approaches such as Markovian Decision Processes (MDPs) have gained attention due to their ability to model traffic dynamics as stochastic processes, enabling more adaptable and efficient signal control strategies. Markovian Decision Processes offer a powerful framework for optimizing traffic signal operations by providing a way to model the uncertainty and variability of traffic conditions at intersections. In an MDP, the decision-making process is based on states, actions, and rewards, where the system transitions from one state to another based on chosen actions, with the objective of maximizing a cumulative reward, such as minimizing delays or maximizing traffic throughput. This approach allows traffic signal systems to continuously adapt to changing traffic volumes, congestion levels, and other environmental factors in real-time.

Despite the potential of MDPs, the complexity of real-time traffic management requires the development of efficient algorithms that can balance short-term decisions with long-term system performance. To tackle this, reinforcement learning (RL) methods can be employed to iteratively improve the control policies, enabling systems to learn optimal strategies without requiring explicit programming for every scenario. This paper explores the application of MDPs and RL for optimizing traffic signal control, aiming to enhance overall traffic flow, reduce congestion, and improve the efficiency of urban transportation systems. Through a series of simulations and experimental evaluations, we demonstrate the effectiveness of the proposed MDP-based approach in comparison to traditional traffic signal control techniques. The results suggest that our framework offers a significant improvement in managing traffic signal timing, with potential benefits for reducing delays, improving throughput, and alleviating congestion in busy urban intersections.

The Markov decision process (MDP), also known as the controlled Markov process, has been extensively explored by researchers since the 1950s, as noted in works such as [1]. It has found applications across a wide range of fields. A discrete-time, stationary Markov control model is characterized by the tuple (X, A, P, R), where: X represents the state space, with each element x ∈ X referred to as a state; A is the set of all possible actions or decisions; P is a probability measure space, with $p_{i,j}^{k}$ indicating the probability of transitioning from state i to state j under action k; and R is a measurable function known as the one-step reward.
Selecting a specific action leads to an immediate reward and a transition probability to the subsequent state. The primary goal is to determine the maximum (least upper bound) of the total expected discounted reward over an infinite time horizon:

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

Where r is the one-step transition reward, $(0 \leq \beta < 1)$) 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:

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

Thus, for a given control problem, once the transition matrix and reward matrix are established, the optimal policy can be derived by maximizing the total expected reward. This policy dictates the best action to take in each state, representing the optimal strategy to follow.
In the subsequent sections, a novel approach grounded in the Markov decision theory outlined above is introduced and applied to the traffic signal control problem.

Traffic Dynamic Model for an Intersection

Modeling traffic flow and optimizing signal control are closely related challenges. A typical four-way intersection is illustrated in Fig. 1. This intersection consists of four approaches, each containing a through movement and a left turn movement, resulting in a total of eight movements. The number associated with each movement is labeled according to the NEMA (National Electrical Manufacturers Association) standard.
Consider a continuous traffic flow process that is sampled at discrete time intervals of duration At, indexed by k. The output of the intersection (i.e., the number of vehicles exiting the intersection) $\underline{q_{out}}(k)$ can be represented as a vector:

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

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

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

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

$$\underline{q_{out}}(k) = \underline{f_{out}}(\underline{u}\,(k), \underline{q}\,(k))$$

Where $\underline{f_{out}}(k)$ is also vector:

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

And

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

Where j = 1, 2, ..., 8. h_{min} is the minimum headway, and $u^{j}(k)$ is the control signal for the j^th movement:
$u^{j}(k) = 1$ denotes a green signal, $u^{j}(k) = 0$ indicates a red signal.

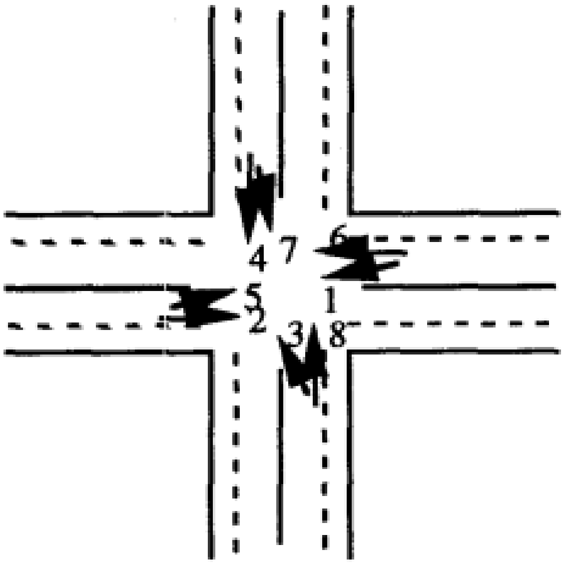

Fig 1: A typical traffic intersection

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 $\underline{q}\,(k-1)$ is the queue at the previous time instant (k-1) and $\underline{q_{in}}\,(k)$ is the input (number of vehicles) during time interval [k- 1, k).
The duration of the current signal, z, must be constrained within a specified minimum and maximum time range:

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

In an eight-phase dual-ring control system, the phases are grouped into two sets (rings) separated by a barrier. Within each ring, four movements (two through movements and their corresponding left turn movements) must be served if there is demand. In theory, there are 2 * 4! = 48 possible phase sequences, but in practice, to prevent conflicting traffic flows, only a subset of sequences (10 out of 48) are permitted (refer to [6] for more details). Since each ring can accommodate up to three permitted phases, the current control decision requires considering the previous three control signals to ensure the sequence constraint is satisfied.

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

Traffic Signal, Control using Markov Decisions

To implement Markovian control in traffic systems, it is necessary to define a state space X and a probability measure P. A threshold (in terms of vehicle count) is selected for the queue of each movement at an intersection. If the queue length for a particular movement exceeds the threshold, that movement is considered to be in a congestion mode; otherwise, it is in a non-congestion mode. These two modes (congestion/non-congestion) are treated as the two states within the state space X. The signal phasing can be regarded as different alternatives for each state. For simplicity, let's assume traffic flow occurs in only two directions: north/south (denoted as 1) or east/west (denoted as 2). In this case, four possible scenarios exist: a) both directions are non-congested; b) direction 1 is congested while direction 2 is non-congested; c) direction 2 is congested while direction 1 is non-congested; and d) both directions are congested. These four scenarios can be defined as the four states of the Markov process. Moreover, in a system with 8 independent movements controlled by an 8-phase signal, the traffic control problem can be modeled as a 256-state Markov process, with 8 alternatives available in each state.

The state space is discrete, so the probability measure P is defined as a discrete transition rule. An element of this matrix, $P^k_{i,j}$, represents the transition probability from state i to state j under alternative k. Let π(k) be a row vector of state probabilities, where $\pi_i(k)$ is the probability that the system will be in state i after k transitions. In the traffic control problem, the probability matrix P is time-varying due to the fluctuating traffic flow, hence:

$$\underline{\pi}(k+1) = \underline{f_\pi}\left[\underline{u}(k), \underline{P}(k)\right]$$

where the probability matrix P(k) depends on the current queue, the predicted number of arrivals in the next time interval, and the control signal:

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

The probability matrix can be defined based on various arrival patterns. In most cases, vehicle arrivals at an isolated intersection follow a Poisson distribution, i.e.:

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

Where n = 1, 2, ..., λ represents the arrival rate, and At is the time interval. Assuming that at a given time, the current queue length for a specific movement i is denoted by q, and q_g vehicles pass through the intersection when the signal for that direction is green, then:

$$P^{u_i}_{S_i\to N_i} = P(\hat{q}^i_{in} + q^i - \delta(u_i)q^i_g \le q^i_{threshold}$$

And

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

Where

$$\delta(u_i) = \begin{cases} 1, where\ u_i = G_i \\ 0, othere\ wise \end{cases}$$

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 identical dimensions and a comparable definition to the probability matrix. The goal in this case is to minimize the queue length, so the queue length functions for various states are selected to form the reward matrix:

$$R^{u_i}_{state1,state2} = f_u(q^i_0, q^i_{threshold}, u_i)$$

Once the transition matrix and reward matrix are determined, an optimal policy for selecting a specific alternative in each state can be derived by maximizing the total expected reward. It has been demonstrated that this optimal solution is unique and can be computed iteratively [1]. Therefore, the problem of selecting signal phasing becomes a decision-making problem for a Markov process.

The adaptive control process for a traffic intersection involves two components: a probability calculation and a Markov decision-making step that utilizes both the probability and the reward (Fig. 2).

The probability matrix and reward matrix are time-varying in the traffic control problem, as they depend on the current state of each traffic movement. Adaptive control requires future arrival information, but accurately predicting long-term arrivals is challenging due to the inherent randomness of the traffic system. Therefore, the sampling frequency should be as high as possible. However, a higher sampling rate increases both the cost and computation time. For this reason, we select $\Delta t = \tau_{min_i}$ (the minimum green extension time). Every At seconds, the time-varying probability matrix P and reward matrix are recalculated, and a decision is made on the control signal for the next time interval based on current measurements from the detector and our estimation. Once the optimal policy is determined, it is implemented for At seconds. In the next time interval, the probability and reward matrices are updated, and the decision-making process is repeated.

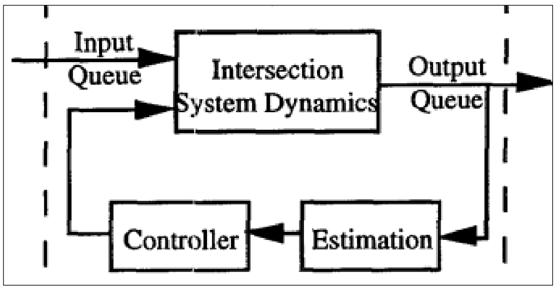

Fig 2: Block diagram of traffic control system for one intersection

Simulation Results

The proposed Markovian adaptive control algorithm was tested through simulations for an isolated intersection, where a Poisson arrival pattern was used as the external input, in order to assess its performance against a fully actuated control method. The following is a summary of some of the simulation parameters:

- 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 evaluated using four different arrival rates: 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 across forty different sets of random data. The mean, covariance, and standard deviation of the average steady-state delay (across the 40 data sets) were calculated and presented in Table 1, where "MAC" refers to the Markov adaptive control algorithm and "FAG" refers to the fully actuated control. Additionally, the means of the steady-state delay for the 40 data sets are plotted in Fig. 3, with the solid line representing the Markov algorithm and the dotted line representing the fully actuated control.

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**
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	26.15
Std.	1.58	2.23	2.64	1.76	4.76	2.03	3.64	2.94	9.61	15.04

Table 2: Bounds for simulation results

	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	4.10	18.75	48.0	91.5	96.67

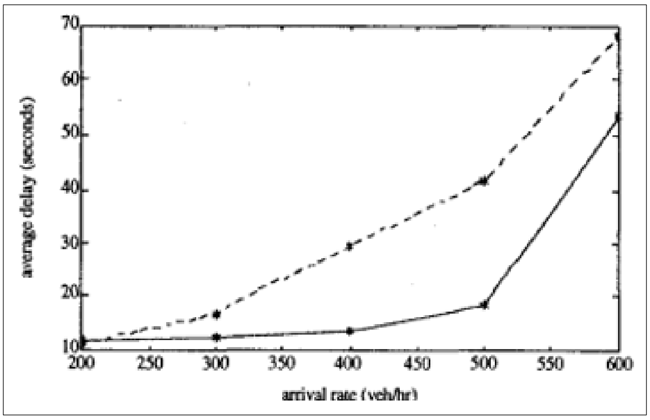

Fig 3: Mean of two algorithms

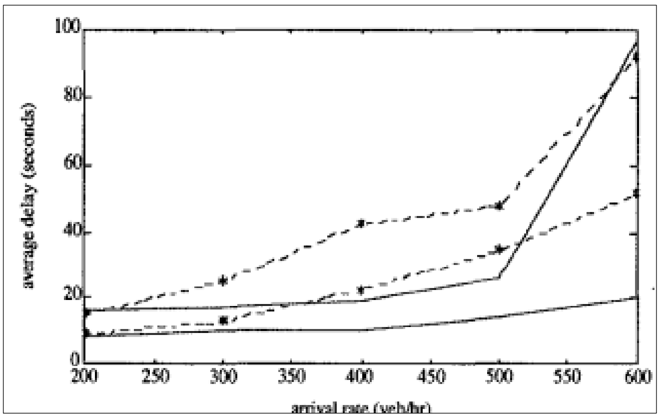

Fig 4: Bounds for simulation results

By applying the concept of distribution-free order statistics, the limits within which at least 90% of the steady-state delay probabilities, as obtained from the simulation, are found with 92% confidence, are shown in Table 2. In other words, there is a 92% probability that 90% of the delay times will fall between the specified lower and upper bounds. These bounds are also illustrated in Figure 4.

The data indicates that when traffic volume is low (e.g., arrival rate of 200 vehicles/hour/movement), the performance of the Markov algorithm is similar to that of the fully actuated controller. However, as traffic volume increases, particularly for arrival rates (λ) of 400 and 500 vehicles/hour/movement, the Markov algorithm outperforms the traditional controller. For instance, at λ = 300, the Markov algorithm improves the average steady-state delay by approximately 26.19%. At λ = 400 and λ = 500, the average steady-state delay with the Markov controller is about half that of the fully actuated controller. As λ continues to increase, the intersection becomes saturated, leading to significant delays for both algorithms.

Conclusions

The traffic system is inherently stochastic. This paper introduces a novel approach to traffic signal control based on Markov decision theory, with computer simulation results and analysis provided. The simulation demonstrates that the proposed approach outperforms the traditional fully actuated control method. Ongoing evaluation and further testing are being conducted.

Acknowledgment

This research was funded by the California Department of Transportation through its Advanced Transportation Management Systems Test Bed Program.

References

1. Hernandez-Lerma O. Adaptive Markov Control Processes. New York: Springer-Verlag; 1989. (Applied Mathematical Sciences; vol 79).
2. Recker W, Ranianathan B, Yu XH, *et al*. Markovian real-time adaptive control of signal systems. J Math Comput Model. [Year];[Volume(Issue)]:[Page range].
3. Transportation Research Center. Traffic Network Study

Tool: TRAATSYT-7F Software Summary. Gainesville: University of Florida; 1987.
4. Hunt PB, Robertson DI, Bretherton RD, *et al*. The SCOOT on-line traffic signal optimization technique. Traffic Eng Control. 1982;23(4):[Page range].
5. Lowrie PR. The Sydney Coordinated Adaptive Traffic System - Principles, Methodology, Algorithms. In: IEE Conference on Road Traffic Signalling; 1982; London. IEE Conference Publication No. 207; 1982. p. [Page range].
6. Wilshire R, Black R, *et al*. Traffic Control Systems Handbook. Washington, DC: Federal Highway Administration; 1985. Report No.: FHWA-IP-85-12.
7. Lu J, Li B, Li H, *et al*. Expansion of city scale, traffic modes, traffic congestion, and air pollution. Cities. 2021;108:102974.
8. Wei H, Zheng G, Gayah V, *et al*. Recent advances in reinforcement learning for traffic signal control: A survey of models and evaluation. ACM SIGKDD Explor Newsl. 2021;22(2):12-18.
9. Marshall WE, Dumbaugh E. Revisiting the relationship between traffic congestion and the economy: A longitudinal examination of US metropolitan areas. Transportation. 2020;47(1):275-314.
10. Afrin T, Yodo N. A survey of road traffic congestion measures towards a sustainable and resilient transportation system. Sustainability. 2020;12(11):4660.
11. De Oliveira LFP, Manera LT, Da Luz PDG. Development of a smart traffic light control system with real-time monitoring. IEEE Internet Things J. 2020;8(5):3384-93.
12. Wang F, Lai G. Fixed-time control design for nonlinear uncertain systems via adaptive method. Syst Control Lett. 2020;140:104704.
13. Eom M, Kim BI. The traffic signal control problem for intersections: A review. Eur Transp Res Rev. 2020;12:50.
14. Wang T, Cao J, Hussain A. Adaptive traffic signal control for large-scale scenario with cooperative group-based multi-agent reinforcement learning. Transp Res Part C Emerg Technol. 2021;125:103046.