Research Article

Optimal Signal Control Design for Isolated Intersections Using Sample Travel-Time Data

HimaBindu Maripini, Lelitha Vanajakshi, and Bhargava Rama Chilukuri

Department of Civil Engineering, Indian Institute of Technology Madras, Chennai, Tamil Nadu 600036, India

Correspondence should be addressed to Lelitha Vanajakshi; lelitha@iitm.ac.in

Received 26 January 2022; Revised 13 April 2022; Accepted 21 April 2022; Published 2 June 2022

Academic Editor: Rui Jiang

Copyright © 2022 HimaBindu Maripini et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Increased travel times are often observed on urban roads, with signalized intersections being the major bottlenecks. The inability of existing static signal timings in accommodating the actual demand fluctuations could be one of the contributing factors. A traffic-responsive signal control system that changes signal timings according to traffic volume fluctuations may alleviate this problem. However, such problems are conventionally formulated based on the data collected from location-based sensors, which are infrastructure intensive and costly and fail to capture mixed and disordered traffic conditions. Considering these limitations, this paper presents an optimal signal design using sample travel time information collected from mobile data sources such as GPS/Bluetooth/Wi-Fi sensors that work independently of the traffic conditions and are relatively cost-effective. The proposed adaptive signal design minimizes total intersection delay at isolated intersections for every cycle based on the traffic conditions observed in the previous cycle. The mathematical programming-based formulation uses shock waves formed during the red and green phases to estimate optimal-phase durations. Results revealed that the proposed design is capable of handling traffic flow fluctuations without requiring the entire traffic stream data. The system demonstrated that sample data from four probe vehicles per phase is adequate for real-time optimal signal design. Results showed that the proposed model outperformed the existing Webster’s signal design procedure with a delay reduction of 11.78% when compared theoretically and 10.41% when implemented in VISSIM.

1. Introduction

Increasing congestion at signalized intersections has always been a concern for road users as well as traffic engineers. It could be partly because of the inability of traffic signal settings to adequately adapt to the dynamic variation in traffic volume. One of the possible reasons for this is the fact that most of the signalized intersections are controlled by fixed-time signal plans, designed based on historical peak period flows [1]. The majority of the conventional signal design procedures assume that a constant arrival flow rate and saturation flow rate adequately represent the prevailing traffic conditions, which may not be valid under highly varying traffic conditions. Consequently, design procedures based on such assumptions lead to suboptimal performance, deteriorating the Level of Service (LoS) of urban networks.

A signal design that can take into account the variations in the demand will be an ideal solution. Signal control systems that can adapt to the traffic flow fluctuations mostly maximize throughput. However, optimizing spatial measures that deal with individual vehicle’s travel experience yield better performance accommodating the variation in traffic dynamics. Delay is one of the most widely used Measures of Effectiveness (MoE) of signalized intersections, as it can directly reflect the motorist’s travel experience. In addition, a majority of the existing studies have chosen control delay to be the performance measure in developing optimal signal control design [2–10]. Hence, this paper proposes a formulation for optimal signal control strategy that minimizes total delay experienced in traversing through an intersection, taking into account the cycle-to-cycle fluctuations in traffic flow.

To achieve the objective of intersection delay minimization, one needs to collect accurate information on the delay experienced by vehicles while traversing through an intersection. Conventional data collection approaches include the use of
inductive loop detectors or video cameras [11]. However, installation and maintenance of loop detectors are laborious and are affected by wear and tear over time, leading to measurement errors [12]. Similarly, video cameras perform unsatisfactory during poor lighting, bad climate, and occlusion [12]. To overcome these limitations, probe vehicles are considered as an alternative source of information. Though the uncertainty in sample size and its ability to represent population using the limited sample data is a concern, it has become a promising source of traffic information in recent times.

Various mobile traffic sensors, such as moving car observers, license plate recognition, GPS (Global Positioning System), RFID tags, and so on, act as probe vehicles and provide travel time information of these vehicles within the traffic stream. With the advances in wireless sensing technologies, use of Media Access Control (MAC) ID matching technique is gaining prominence for traffic applications, with Bluetooth and Wi-Fi MAC scanners as sources of travel time information [13–16]. Two such scanners placed at upstream (beyond queue position) and downstream locations of an intersection can be used to get the travel time of vehicles [17, 18]. Hence, Wi-Fi MAC Scanners have been considered a reliable traffic data collection source for travel time and thereby delay [19–22]. Therefore, estimation of total intersection delay from any of the probe data sources of minimum penetration rate is the major objective to be accomplished.

Existing literature on delay estimation uses the concept of delay being the area under the queue polygon [23]. Most of the studies used input-output analysis of queue formation and dissipation processes. They assume that the queue accumulates in the form of vertical stacks with density being constant throughout the queue formation and dissipation processes [24]. Such an implicit assumption fails to capture the traffic dynamics in the vicinity of the intersection and becomes crucial at near-saturation conditions [25]. To address this, Shock Wave (SW) theory based on the LWR model [26, 27] is applied to explain the queue formation and dissipation processes [28].

Shock wave is a macroscopic representation of traffic behavior with the change in traffic states and is represented with the help of space-time plots (x-t plots). Though this involves the assumption of instantaneous acceleration and deceleration of vehicles, it has wider applicability in analyzing traffic dynamics in both undersaturated and oversaturated conditions. Since SW-based models are derived from fundamental traffic flow concepts, they are considered to be theoretically sound and computationally efficient [29]. Therefore, the present study incorporates the SW theory to estimate total intersection delay using travel time information from Wi-Fi sensors and develop an optimal traffic control strategy for isolated intersections. The major contributions made in this process are as follows:

(i) Use of robust and cost-effective mobile data sources for traffic control applications
(ii) Development of an adaptive traffic control system using sparse data
(iii) A methodology to estimate various traffic state variables, such as delay, queue length, and arrival flow rate using sample data
(iv) An optimal signal design that can be implemented in real time for mixed traffic conditions.

The remainder of this paper is structured as follows. Section 2 presents a review of the literature on probe based traffic signal control. Section 3 discusses the problem formulation adopted in this study. Section 4 presents the implementation of the proposed model with significant findings, followed by performance comparison with the existing signal design procedures.

2. Literature Review

This section presents a review of relevant literature on probe-based traffic signal control. The last part of this section highlights the gaps in the literature and contributions of this paper.

The use of probe data in traffic engineering applications has evolved over the last two decades in the field of intelligent transportation systems. Guo et al. [29] classified various approaches to traffic flow modelling using probe data into (i) stochastic learning approach, (ii) shock wave-based approach, and (iii) kinematic equation-based approach. Various applications of these approaches include queue length estimation, travel time estimation, traffic state estimation, traffic safety, traffic signal control, and reconstruction of vehicle trajectories [30]. These studies used mobile sensors, such as GPS- (Global Positioning System-) fitted vehicles [31–34], a combination of both probe- and location-based sensors [35, 36], and connected vehicle technologies [29, 37]. From these, studies related to traffic signal control using data from probe vehicles are discussed below.

Probe data was initially used for performance evaluation in traffic signal control rather than for design purposes. To achieve signal coordination, Day et al. [38] introduced a quantitative performance measure that relates the number of vehicle arrivals on green and the associated offsets. Data used to evaluate the performance measure was obtained from Bluetooth devices by MAC ID matching technique. Similarly, Hunter et al. [39] evaluated the performance of the adaptive traffic control system, SCATS based on speed and travel time data collected from Global Positioning System- (GPS-) instrumented test vehicles. Shladover and Li [40] evaluated the efficiency of the probe data in estimating MoEs of signal control in comparison to the field values, using simulation-based probe sampling strategies. It was inferred from the study that either aggregating probe data over multiple cycles or fusing the data from fixed detectors might lead to reliable traffic estimates in real time. Use of probe data in signal design was reported by Chen et al. [41], coupled with location-based sensors to minimize the waiting time of vehicles. Much later, Hu et al. [42] developed a procedure to coordinate urban arterials using event-based data. Nagashima et al. [43] developed an optimal signal control methodology related to origin-destination
adaptation and determined the optimal splits and cycle length values based on the traffic volume entering the intersection and the queue length formed. The required queue length information was estimated with the help of GPS data and detector data. Park and Haghani [44] used Bluetooth (BT) data for arterial signal coordination. Link-based trajectories of vehicle platoons were obtained from BT data and linked with signal timings such that maximum bandwidth progression is achieved. Using sparse data, a multivariate time series model was developed by Moghimi et al. [45] to achieve a fully actuated signal control. Wang et al. [46] proposed coordinated traffic control model for multiple coupled intersections based on vehicle trajectories. A linear integer programming formulation is presented for joint optimization of vehicle space-time trajectories and traffic control. The proposed approach was shown to perform better in coordinating vehicles’ trajectories under various traffic conditions. It can be inferred from the above optimal signal strategies that the majority of them used probe data to achieve signal coordination across intersections.

Probe data-based adaptive optimal signal design studies are reported in the recent times and are limited. Ezawa and Mukai [47] proposed adaptive traffic signal control system based on the position and path information obtained from the route sharing of probe vehicles. This information is used for calculating the criteria of traffic congestion called expected traffic congestion and the optimal traffic signal parameters that minimize the expected congestion are obtained. Lian et al. [48] proposed two new adaptive traffic signal control algorithms for signal coordination on arterial roads. While the iterative adaptive control algorithm assigns green to the links according to the proportion of moving vehicles estimated from probe data, the optimized control algorithm minimizes the product of the proportion of moving vehicles’ weightage and the green time difference between upstream and downstream intersections. The results revealed a substantial reduction in travel times, delays, and number of stops. Yao et al. [49] proposed a dynamic optimization method for adaptive signal control based on connected vehicles, where a dynamic platoon dispersion model is developed to predict the vehicle arrivals. From the predicted arrivals, average vehicle delay is estimated and optimized to obtain optimal signal timing plan. It was shown from the results that the model performs well under 100% penetration rate and the performance exponentially decreases with decrease in penetration rate. From this, it can be inferred that the above referred studies majorly focused on optimal signal coordination than the optimal performance of individual intersections at local level.

Some of the optimal traffic signal control strategies based on shock wave theory to improve intersection performance are discussed next. The earliest study on the application of shock wave theory was reported by Michalopoulos [24], in which minimization of total intersection delay subjected to queue length constraints was studied and was shown to augment the performance of both undersaturated and oversaturated intersections. Much later, Ponlathep [50] developed a simple adaptive traffic signal control algorithm to obtain green splits and cycle length values that minimize total intersection delay and residual queue lengths, respectively. The derivation of the performance measures was based on shock wave parameters estimated from the occupancy data of loop detectors. Noaeen et al. [51] proposed an optimization model that minimizes the shock wave delay, expressed as a function of lost time of each of the phases. Recently, Mohajerpoor et al. [52] presented an analytical framework to minimize delay variability and spillback probability, in addition to the total intersection delay, and obtained optimal split and cycle length values. However, these shock wave theory-based studies require the information on the traffic demand arriving at the intersection for analyzing the queue formation and dissipation processes. This arrival flow information is generally collected with the help of location-based sensors. Ban et al. [53] identified the pattern of intersection delays using probe data under uniform vehicle arrivals with known signal information. A least square-based linear fitting algorithm was used to estimate intersection delay. This work was later extended to estimate various traffic state variables, such as queue lengths, at signalized intersections [54]. However, estimating the arrival flow from probe data for optimal signal design is a challenging task.

From the literature review presented above, it is evident that very few studies used probe data to obtain optimal signal settings of signalized intersections and even those studies either supplemented the probe data with location-based sensor data for traffic control or focused on signal coordination. A dynamic signal control system that depends solely on probe data and developed based on shock wave theory is not reported. For the first time, this study proposes to use limited sample data from a cost-effective data source for real-time traffic control. With the emergence of Bluetooth and Wi-Fi sensors that are reliable, portable, easy to install and maintain on roadside, cost-effective, and collect sufficient amount of probe data in real time, they may be used to adequately capture the traffic variations in real time for traffic signal control applications. Towards this goal, this study presents a dynamic signal control methodology based on sample travel time information. This methodology can enable easy upgrade of the existing fixed-time signals to traffic responsive without the need for large-scale installation of expensive sensors. In addition, the proposed algorithm ensures queue dissipation and thereby undersaturation on the approaches.

3. Model Formulation

Consider an isolated intersection with \( P \) phases and a constant cycle length, with two probe data sources placed at upstream and downstream of the intersection \( L \) kilometers apart. If \( t_i \left( f_i \right) \) and \( t_{i+1} \left( f_{i+1} \right) \) are upstream (downstream) time stamps of two vehicles, \( i \) and \( i + 1 \), detected by any mobile traffic sensor, their travel times are obtained as the differences in their respective time stamps. Consequently, corresponding delay values \( d_i \) and \( d_{i+1} \) can be obtained as the
extra time spent compared to free-flow traffic conditions, using equation (1). If all such delay values experienced by vehicles while traversing through an intersection are plotted with respect to their corresponding vehicle arrivals at the downstream location, it is defined as a delay polygon, as shown in Figure 1.

\[ d_p^i(m) = \left( t_p^i(m) - t_p^i(m) \right) - L_p^u - L_p^f. \]  

The objective of the formulation is to minimize the sum of areas of delay polygons of all \( P \) phases of the intersection to obtain optimal signal timings for the phases of the cycle. Since the slopes of lines joining delay values of probe vehicles are piecewise linear, the variation in delay experienced by vehicles arriving between successive probe vehicles and between actual and probe vehicles changes linearly throughout the cycle. Accordingly, piecewise shock wave speeds between arrivals of consecutive probe vehicles are approximated to an average constant shock wave speed throughout the cycle. The number of vehicles that reach the downstream location at saturation flow rate, between the first and last vehicles, \( (n_p^f) \) and \( (n_p^i) \), can be written as

\[ (n_p^f(m) - n_p^i(m)) = S_p \left( t_p^f(m) - t_p^i(m) \right). \]

Substituting (4) and (5) in (3), the objective function can be expressed as

\[ \text{Min} \sum_{p} A_p^f(m). \]
Table 1: List of notations.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p )</td>
<td>Phase index, of range ([1, 2, \ldots, P])</td>
</tr>
<tr>
<td>( m )</td>
<td>Cycle index, of range ([1, 2, \ldots, M])</td>
</tr>
<tr>
<td>( i )</td>
<td>Index of vehicles detected by mobile data source during the ( p^{th} ) phase of ( m^{th} ) cycle, of range ([1, 2, \ldots, 1^{P}(m)])</td>
</tr>
<tr>
<td>( n_f )</td>
<td>Number of the first vehicle at the downstream location that arrives at the start of the red phase</td>
</tr>
<tr>
<td>( n_l )</td>
<td>Number of the last vehicle at the downstream location that is affected by red phase</td>
</tr>
<tr>
<td>( T^p(m) )</td>
<td>Time taken to dissipate the queue formed during red of ( p^{th} ) phase of ( m^{th} ) cycle in hours</td>
</tr>
<tr>
<td>( t^p_i(m) )</td>
<td>Time stamp at which ( k^{th} ) vehicle of ( p^{th} ) phase during ( m^{th} ) cycle passes the upstream location</td>
</tr>
<tr>
<td>( \bar{T^p}(m) )</td>
<td>Time stamp at which ( k^{th} ) vehicle of ( p^{th} ) phase during ( m^{th} ) cycle crosses the downstream location</td>
</tr>
<tr>
<td>( d^p_i(m) )</td>
<td>Delay experienced by ( k^{th} ) vehicle of ( p^{th} ) phase of ( m^{th} ) cycle in hours</td>
</tr>
<tr>
<td>( z^p_i(m) )</td>
<td>Jam spacing between ( k^{th} ) and ((k+1)^{th}) probe vehicles of ( p^{th} ) phase of ( m^{th} ) cycle in km</td>
</tr>
<tr>
<td>( z^{\text{max}}_i(m) )</td>
<td>Maximum queue length from the stop line during queue formation of ( p^{th} ) phase of ( m^{th} ) cycle in km</td>
</tr>
<tr>
<td>( n^M_i(m) )</td>
<td>Number of ( k^{th} ) vehicle on arrival at the downstream location in ( p^{th} ) phase of ( m^{th} ) cycle</td>
</tr>
<tr>
<td>( s^p_i(m) )</td>
<td>Actual backward queue forming shock wave speed of ( p^{th} ) phase of ( m^{th} ) cycle in kmph</td>
</tr>
<tr>
<td>( s^{\text{Estimated}}_i(m) )</td>
<td>Estimated backward queue forming shock wave speed between ( k^{th} ) and ((k+1)^{th}) vehicles of ( p^{th} ) phase of ( m^{th} ) cycle in kmph</td>
</tr>
<tr>
<td>( q^p_i(m) )</td>
<td>Arrival flow rate of ( p^{th} ) phase during ( m^{th} ) cycle in veh/hr</td>
</tr>
<tr>
<td>( \varphi^p(m) )</td>
<td>Capacity of ( p^{th} ) phase during ( m^{th} ) cycle in veh/hr</td>
</tr>
<tr>
<td>( r^\text{max} )</td>
<td>Upper bound of effective red duration of ( p^{th} ) phase during ( m^{th} ) cycle in hours</td>
</tr>
<tr>
<td>( r^\text{min} )</td>
<td>Lower bound of effective red duration of ( p^{th} ) phase during ( m^{th} ) cycle in hours</td>
</tr>
<tr>
<td>( g_{\text{max}} )</td>
<td>Upper bound of effective green duration of ( p^{th} ) phase during ( m^{th} ) cycle in hours</td>
</tr>
<tr>
<td>( g_{\text{min}} )</td>
<td>Lower bound of effective green duration of ( p^{th} ) phase during ( m^{th} ) cycle in hours</td>
</tr>
<tr>
<td>( u^p_r )</td>
<td>Free-flow speed of vehicles in kmph</td>
</tr>
<tr>
<td>( \rho^p_r )</td>
<td>Critical density in veh/km</td>
</tr>
<tr>
<td>( \omega^p )</td>
<td>Wave speed in kmph</td>
</tr>
<tr>
<td>( \rho^p )</td>
<td>Jam density in veh/km</td>
</tr>
<tr>
<td>( \rho^p )</td>
<td>Saturation flow rate in veh/hr</td>
</tr>
<tr>
<td>( C )</td>
<td>Cycle length in hours</td>
</tr>
<tr>
<td>( L^p )</td>
<td>Distance between upstream and downstream detector locations corresponding to ( p^{th} ) phase, in km</td>
</tr>
<tr>
<td>( T^p )</td>
<td>Average pedestrian crossing time for ( p^{th} ) phase, in hours</td>
</tr>
<tr>
<td>( r^p(m) )</td>
<td>Optimal effective red duration of ( p^{th} ) phase of ( m^{th} ) cycle, in hours</td>
</tr>
</tbody>
</table>

\[
\sum_{p} A^p(m) = \sum_{p} \left( \frac{1}{2} r^p(m) S^p\left( t^p_i(m) - t^p_{i-1}(m) \right) \right). 
\] (6)

It is to be noted that (6) comprises of unknown parameters related to effective red duration \((r^p(m))\) and queue dissipation time \((t^p_i(m) - t^p_{i-1}(m))\). To reduce the problem as a function of effective red duration \((r^p(m))\) only, the queue dissipation time is derived based on shock wave theory.

Shock wave theory explains the traffic dynamics resulting from the changes in signal timings near an intersection. Shock wave diagram of queue formation and dissipation processes during red is explained with the help of space-time \((x - t)\) plots of Figure 3(a) and the triangular fundamental diagram shown in Figure 3(b). A represents free-flow arrival state with flow rate and density as \(q_A\) and \(\rho_A\), respectively, and B represents the queue state during red with \(\rho_f\) as the jam density, shock wave speed generated while queue formation between traffic states A and B is expressed [55] as

\[
s = \frac{q_A}{(\rho_A - \rho_f)}. \] (7)

From (7), it can be observed that the shock wave speed is dependent on flow and density, which can be challenging and resource intensive to measure using traditional location-based sensors, such as loop detectors. Therefore, an alternative approach to estimate the shock wave speed using probe vehicle data is developed as explained below.

The average shock wave speed \((s^p_i(m))\) during queue formation is obtained as the weighted average of individual shock wave speeds of each consecutive probe vehicle pair, \(i\) and \(i + 1\), as

\[
s^p_i(m) = \frac{\sum_{i=1}^{M} s^p_i(m) \left( t^p_{i+1}(m) - t^p_i(m) \right)}{\left( t^p_i(m) - t^p_{i-1}(m) \right)}. \] (8)

Individual shock wave speed between any two consecutive probe vehicle arrivals \((s^p_i(m))\) can be obtained from the travel time data (i.e., time stamps of probe vehicles at upstream and downstream locations are related to queue lengths from the “zoomed-in” sections of Figures 3(c) and 3(d)) as

\[
s^p_i(m) = \frac{t^p_{i+1}(m) - t^p_i(m)}{\left( 1/\omega^p \right) + \left( 1/u^p \right)}. \] (9)

From this, the shock wave speed between the arrivals of \(i\) and \(i + 1\) probe vehicles, \(s^p_i(m)\) can be derived as
The queue dissipation time \( X_p(m) \) (i.e., the time taken between the arrival of last and first vehicles at the downstream location \((\bar{t}_n^p(m) - \bar{t}_f^p(m))\)) can be expressed in terms of the maximum length \( z_{\text{max}}^p(m) \) as

\[
(\bar{t}_n^p(m) - \bar{t}_f^p(m)) = z_{\text{max}}^p(m) \left( \frac{1}{\omega^p} + 1/u_p^p \right).
\]

(11)

Maximum queue length \( z_{\text{max}}^p(m) \) is a function of effective red duration and the uniform shock wave speed from Figure 3(a) that can be expressed as

\[
z_{\text{max}}^p(m) = \frac{(r_p(m))}{(1/s^p(m) - 1/\omega^p)}.
\]

(12)

Substituting (11) and (12), the objective function in (6) can be simplified as

\[
\sum_p A_p^p(m) = \sum_p \left( \frac{1}{2} S_p^p \left( \frac{1/\omega^p + 1/u_p^p}{1/s^p(m) - 1/\omega^p} \right) (r_p(m))^2 \right). \]

(13)

Among the unknown parameters of (13), the estimated shock wave speed \( s^p(m) \) can be obtained from (8) and (10). The only unknown independent variable being effective red duration \( r_p(m) \) and for being a

major constituent of delay, effective red duration of each phase is considered as the decision variable of the objective function. The associated constraints are presented below.

1. Effective red durations of all phases are nonnegative:

\[ r_p(m) \geq 0, \quad \forall p, \]

(14)

2. Sum of the red durations of all phases is an integral multiple of cycle length:

\[ \sum_p r_p(m) = (P - 1) \cdot C + L. \]

(15)

3. Effective red duration is constrained by lower and upper bounds:

\[ r_{\text{min}}^p(m) \leq r_p(m) \leq r_{\text{max}}^p(m). \]

(16)

The upper bound of red duration can be expressed as

\[ r_{\text{max}}^p(m) = C - g_{\text{min}}^p(m). \]

(17)

The minimum effective green is taken as the maximum of queue dissipation time, \( X_p(m) \), to be allotted in order to ensure undersaturation, and pedestrian crossing time \( T_P \), which can be expressed as

\[ g_{\text{min}}^p(m) = \max(X_p(m), T_P). \]

(18)

Queue dissipation time \( X_p(m) \) can be expressed from (11) and (12) as
Information collected from location-based sensors as this case, it is obtained in terms of flow rate and density shockwave speed when 100% population data is available. In order to yield optimal red duration for each phase of green duration as phase of Model II. The objective function for this case is known from location-based sensors) is considered and a similar model is formulated inclusive of the corresponding yellow time of the phase. Details be noted that the effective red duration under consideration is the lost time of each phase implicitly. In addition, it is to be noted that the effective red duration under consideration is inclusive of the corresponding yellow time of the phase. Details on the type of optimization problem and the consequent solutions obtained are discussed in the implementation section.

To evaluate the performance of this model (hereafter referred to as Model I), a best-case scenario where entire traffic stream data in terms of flow rate and density (known from location-based sensors) is considered and a similar model is formulated (referred to as Model II). The objective function for this case is similar to that of the previous case, which is obtained as the sum of areas of delay polygons of all phases, and is expressed as

\[ X^P (m) = t^P_{e} (m) - t^P_{e} (m) \]

\[ = \frac{(r^P (m))}{1/s^P (m) - 1/\omega^{p}} \left( \frac{1}{\omega^{p} + 1/u^{P}_{i}} \right). \] (19)

The lower bound of red duration can be expressed in terms of green duration as

\[ r^P_{min} (m) = C - g^P_{max} (m) = \sum_{i \in p} s^f_{min} (m). \] (20)

Objective function expressed in (13) subject to the constraints (14)–(20) is minimized by performing optimization to yield optimal red duration for each phase of cycle. The effective red and green allotments considered in the formulation include the lost time of each phase implicitly. In addition, it is to be noted that the effective red duration under consideration is inclusive of the corresponding yellow time of the phase. Details on the type of optimization problem and the consequent solutions obtained are discussed in the implementation section.

To evaluate the performance of this model (hereafter referred to as Model I), a best-case scenario where entire traffic stream data in terms of flow rate and density (known from location-based sensors) is considered and a similar model is formulated (referred to as Model II). The objective function for this case is similar to that of the previous case, which is obtained as the sum of areas of delay polygons of all phases, and is expressed as

\[ \min \sum_{p} A^P (m) = \sum_{p} \left( \frac{1}{2} r^P (m) S^P \left( \frac{\omega^{p} / u^{P}_{i} + 1}{\omega^{p} / s^{P} (m) - 1} \right) \right). \] (21)

The only difference from (13) is the estimation of average shock wave speed when 100% population data is available. In this case, it is obtained in terms of flow rate and density information collected from location-based sensors as

\[ s^P (m) = \frac{q^P (m)}{\rho^P f \left( \frac{q^P (m)}{u^P} \right)} . \] (22)

The decision variable of Model II problem is similar to that of Model I (i.e., the optimal effective red duration of \( p \)th phase of \( m \)th cycle \( r^P (m) \)). Therefore, the set of constraints considered for Model I (equations (14)–(20)) holds good for finding feasible space of Model II. Solving (21) yields optimal red duration \( r^P (m) \) value. Table 2 summarizes the objective function and constraints of the formulation proposed in Model I. However, for Model II, the estimated average shock wave speed \( s^P (m) \) is replaced with the measured shock wave speed \( s^f (m) \) from location-based sensors.

4. Implementation and Results

The proposed methodology is implemented in a micro-simulation environment, VISSION, to evaluate its performance. An isolated intersection of a major and minor road in E-W and N-S directions, respectively, with exclusive turn lanes is simulated in VISSION. It is assumed that the traffic signal has four phases with a split phasing sequence, as shown in Figure 4. To represent realistic traffic flow variations across cycles, arrival flow rates from field are collected for one of the approaches and the variation is considered to be valid for all the phases in the simulation, as shown in Figure 5. The through movements of east and west bound phases and the right turning traffic of north and south bound phases are considered as critical lane movements. Default parameter values of VISSION are used except the desired speed (80 km/hr) and speed distribution bounds (75 km/hr and 85 km/hr, resp.).

To start with, signal timings are calculated based on Webster’s method. The peak flow rates (of cycle 15, as shown in Figure 5) with lost time of 4.0 sec/phase for E-W direction and 3.0 sec/phase for N-S direction are considered for the design. A saturation headway of 1.2 sec/veh is used, based on field observation, for all phases. The optimal cycle length for a cycle with peak traffic volume (Cycle 15) obtained based on Webster’s design is 204 sec (rounded to 210 sec) and is assumed to be constant over all cycles.

Two data collection detectors (equivalent to Wi-Fi scanners) are placed at upstream and downstream locations of each approach to capture the respective travel time of vehicles. Simulation is run for a duration of 3600 seconds, with 18 cycles in total (constant cycle length of 210 seconds). Basic traffic flow parameters obtained from VISSION, associated with a flow-density relationship, are shown in Table 3.

The number of probe vehicles to be considered to obtain travel time information is determined from field observation. It is observed that three to four vehicles on an average are being detected every minute by the Wi-Fi or Bluetooth MAC

| Table 2: Problem formulation of Model I. |
|------------------------------|------------------|
| **Objective function**     | Min. \( \sum_{p} A^P (m) = \sum_{p} \left( \frac{1}{2} r^P (m) S^P \left( \frac{\omega^{p} / u^{P}_{i} + 1}{\omega^{p} / s^{P} (m) - 1} \right) \right) \) |
| **Constraints**            | \( r^P (m) \geq 0, \quad \forall p \) |

The set of constraints considered for Model I (equations (14)–(20)) holds good for finding feasible space of Model II. Solving (21) yields optimal red duration \( r^P (m) \) value. Table 2 summarizes the objective function and constraints of the formulation proposed in Model I. However, for Model II, the estimated average shock wave speed \( s^P (m) \) is replaced with the measured shock wave speed \( s^f (m) \) from location-based sensors.

4. Implementation and Results

The proposed methodology is implemented in a micro-simulation environment, VISSION, to evaluate its performance. An isolated intersection of a major and minor road in E-W and N-S directions, respectively, with exclusive turn lanes is simulated in VISSION. It is assumed that the traffic signal has four phases with a split phasing sequence, as shown in Figure 4. To represent realistic traffic flow variations across cycles, arrival flow rates from field are collected for one of the approaches and the variation is considered to be valid for all the phases in the simulation, as shown in Figure 5. The through movements of east and west bound phases and the right turning traffic of north and south bound phases are considered as critical lane movements. Default parameter values of VISSION are used except the desired speed (80 km/hr) and speed distribution bounds (75 km/hr and 85 km/hr, resp.).

To start with, signal timings are calculated based on Webster’s method. The peak flow rates (of cycle 15, as shown in Figure 5) with lost time of 4.0 sec/phase for E-W direction and 3.0 sec/phase for N-S direction are considered for the design. A saturation headway of 1.2 sec/veh is used, based on field observation, for all phases. The optimal cycle length for a cycle with peak traffic volume (Cycle 15) obtained based on Webster’s design is 204 sec (rounded to 210 sec) and is assumed to be constant over all cycles.

Two data collection detectors (equivalent to Wi-Fi scanners) are placed at upstream and downstream locations of each approach to capture the respective travel time of vehicles. Simulation is run for a duration of 3600 seconds, with 18 cycles in total (constant cycle length of 210 seconds). Basic traffic flow parameters obtained from VISSION, associated with a flow-density relationship, are shown in Table 3.

The number of probe vehicles to be considered to obtain travel time information is determined from field observation. It is observed that three to four vehicles on an average are being detected every minute by the Wi-Fi or Bluetooth MAC
Scanners. Hence, travel times of a minimum of four-probe vehicles per cycle that are uniformly distributed during the green duration are considered to be representative of the entire population and are extracted from the output files of VISSIM.

In order to determine the extent to which the collected sample data represents the entire traffic stream, the estimated shock wave speed values \( s^p(m) \) obtained from travel time of probe vehicles ((11) and (13)) are compared with the actual shock wave speed values \( s^p(m) \) obtained from population data ((22)). Sample plots showing these values for all four phases are shown in Figure 6.

From Figure 6, it can be observed that the average absolute error in estimating shock wave speeds using proposed method is around 1.0 km/hr. Root Mean Square Error (RMSE) over all cycles is evaluated using

\[
RMSE = \sqrt{\frac{\sum (\tilde{y}_i - y_i)^2}{M}},
\]

where \( \tilde{y}_i \) is the mean value of the population data and \( y_i \) is the observation from sample data with \( M \) being the total number of cycles considered. Average RMSE over all phases was found to be 0.49, which implies that the sampled probe vehicles are able to closely represent the whole traffic stream and thus feasible for implementation.

After acquiring the required input data, implementation using an appropriate optimization software is the next step. Since the objective function is quadratic in the decision variable, a quadratic constrained optimization is performed using General Algebraic Modelling System (GAMS) software. GAMS is a high-level modelling tool designed for solving linear, nonlinear, and mixed optimization problems. The present study implemented the formulation in GAMS 23.9 and solved using CPLEX 12.1 solver. The objective function ((13) subjected to a list of linear and nonnegative constraints (equations (14)–(20)), and the input parameters in Table 3 are fed into the GAMS software. It is to be noted that, in cases where \( \sum_{l \neq p} G_{l,i}(m) \) exceeds the cycle length, the cycle length value can be readjusted to \( \sum_{l \neq p} G_{l,i}(m) \). Since the optimal signal timing parameters are to be estimated and implemented for the current cycle \( (m^{th} \) cycle), delay polygons are generated using the sample vehicles of previous cycle that reach the downstream location before the start of the current cycle. Therefore, while implementing the model, the first two cycles of the analysis period are designed based on Webster’s method and the proposed model is implemented for the rest of the cycles.

### Table 3: Input parameters for implementation.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>( u_f ) in kmph</td>
<td>80</td>
</tr>
<tr>
<td>( \omega ) in kmph</td>
<td>15.59</td>
</tr>
<tr>
<td>( S ) (saturation flow rate) in veh/hr/lane</td>
<td>3000</td>
</tr>
<tr>
<td>( \rho_j ) (jam density) in veh/km/lane</td>
<td>230</td>
</tr>
</tbody>
</table>

Figure 4: Phase diagram of the subject intersection.

Figure 5: Variation of flow rate of critical phases of the major corridor.
Since interfacing with VISSIM to send and receive data in real time is possible using the VISSIM COM interface, the proposed system is implemented in real time in VISSIM [56]. The model is programmed in Python coding software and is interfaced with VISSIM through the COM interface. The time stamps of sample probe vehicles (from any phase) at the study intersection are collected in real time using COM interface. The sample travel data is used for the total control delay estimation, which is minimized in the optimization module to calculate the optimal phase durations.
The obtained optimal signal timings are communicated to VISSIM signal control module to be implemented in the next cycle in VISSIM and the process repeats for every cycle providing a dynamic signal control system. Figure 7 shows the implementation details in a flowchart form.

Optimal signal timings obtained are plotted with respect to the variation in flow rate in Figure 8 to verify the adaptability of the model in accommodating varying demand. Green splits obtained from Webster’s method are also included for comparison. It can be seen that the effective green duration is allocated dynamically based on the demand fluctuations by the proposed design, whereas Webster’s signal timings are unable to meet the rising traffic demand of major approaches (Figure 8(a)). Also, it can be observed from the figure that east bound is subjected to a rapid increase in flow rate to peak values in six out of sixteen cycles (cycles: 6, 8, 10, 12, 15, and 17). For the same set of cycles, both north and south bound phases are operating under decreased traffic demand (Figures 8(c) and 8(d)). Still, they are given extra green as per Webster’s signal design, leading to inefficient use of green. On the other hand, the proposed model allots minimum green time to minor phases and prioritizes the traffic flow fluctuations in major phases, improving the total intersection performance.

This is further illustrated in Figure 9, where the variation of green splits (g/C) with the critical flow ratio (q/S) of each phase for the proposed and Webster’s design is plotted. Figure 9(a) shows that the green splits of each phase are allotted in proportion to the variation of critical flow rate and the saturation flow rate of each phase. On the contrary, it can be observed from Figure 9(b) that the green splits are allotted based on the proportion of critical flow ratio of a phase to the sum of critical flow ratios of all phases, irrespective of the variation in the value of saturation flow rate. This reflects the adaptive behavior of the proposed method relative to Webster’s method under traffic flow fluctuations.

4.1. Performance Evaluation. In order to evaluate the efficiency of the sample data relative to population data, Model II is implemented following the same procedure as Model I, subjected to similar traffic conditions. The necessary...
information is collected from location-based detectors placed at upstream and downstream locations of the intersection in VISSIM. Considering Model II as the best-case scenario, the optimal signal timings are obtained for all cycles from GAMS. The corresponding total intersection delay incurred from Model I (13) and Model II (22) is considered as a performance measure and is compared. For a given arrival flow rate, the intersection delay values incurred due to the respective optimal signal timings of Model I and Model II are shown in Figure 10. With respect to the model using 100% population data (Model II), the sample-based model (Model I) showed a difference of only 3.94%. From this, it can be inferred that the proposed design, based on sample data, performed comparable to the one based on entire traffic stream data.

In addition, the performance of optimal signal design is compared with the existing signal design procedure adopted in field (i.e., Webster’s signal design). According to Webster [1], the total delay experienced on each phase can be expressed as

\[
Total \ Delay = 0.9 \left( Uniform \ Delay + Random \ Delay \right). \quad (24)
\]

Uniform Delay component (UD) of phase “p” can be expressed as

\[
UD^p = \frac{1}{2} C^2 \left( 1 - \frac{g^p}{C} \right) \frac{q^p}{(1 - q^p/C^p)} \quad (25)
\]

Random Delay component (RD) of phase “p” can be expressed as

\[
RD^p = \frac{(q^p/C^p)^2}{2q^p(1 - q^p/C^p)} \quad (26)
\]

Phase-wise delays under Webster’s signal timings are computed using equations (24)–(26) and summed up over all phases to obtain intersection delay value for every cycle. The percentage reduction in intersection delay incurred based on sample data based optimal design relative to the
delay values computed using equations (24)–(26) is plotted in Figure 11. It resulted in an average delay reduction of 11.78% over all cycles by implementing optimal signal design. In addition, a left-tailed $t$-test is conducted to verify whether the mean of the theoretical intersection delays of optimal design ($x_O$) are lower than that of the corresponding mean of Webster’s design ($x_W$). With null hypothesis of $H_0: x_O \geq x_W$ and an alternate hypothesis of $H_1: x_O < x_W$, $t$-test is performed at 95% confidence level. The calculated $t$-value ($t_C = -2.34$) is within the rejection region of the critical $t$-value ($t_{critical} = -1.753$). Therefore, it can be concluded that the intersection delay values theoretically obtained from the optimal signal design are lower than those of Webster’s design.

Another evaluation conducted is by implementing signal timings obtained from Model I and Webster’s design in VISSIM. Fifteen scenarios with different random seed values are considered in VISSIM and simulation runs are performed to all 16 cycles (3 to 18 cycles) under Webster’s design and the proposed optimal design. The corresponding percentage reduction in total intersection delay averaged over all cycles is depicted for each of the 15 scenarios in Figure 12 below. From Figure 12, it can be observed that all the scenarios exhibit delay reduction under proposed optimal signal design, with an average reduction of 10.41%. A left-tailed $t$-test is conducted to verify whether the mean of the total intersection delays of optimal design ($x_O$) is lower than that of the corresponding mean of Webster’s design ($x_W$). With null hypothesis of $H_0: x_O \geq x_W$ and an alternate hypothesis of $H_1: x_O < x_W$, a $t$-test is performed at 95% confidence level. The calculated $t$-value ($t_{cal} = -5.04$) is within the rejection region of the critical $t$-value ($t_{critical} = -1.753$). This implies that the intersection delay values obtained from VISSIM of optimal signal design are lower than that of Webster’s design. From this, it can be inferred that the proposed model is far more reliable than
the existing field control procedures, without collecting the whole traffic stream data.

The consistency in the performance of the proposed model with variation in the pattern and the magnitude of traffic demand for five consecutive cycles (cycles 2, 3, 4, 5, and 6) out of the 18 cycles is verified next. The percent reduction in the performance measures, such as intersection delay and queue length of each cycle, is considered and the individual values are averaged over the four random seed runs for each cycle, as shown in Table 4. The standard deviations in the individual values of each of the random seed scenario with respect to the corresponding average values are compared for each cycle and it is found that the values span within the range of 2.1% to 5.2%. When the standard deviation values of all the cycles are averaged, they yielded smaller values, as low as 4.6% and 4.8% for delay reduction and queue length reduction, respectively. Therefore, the performance of the model is inferred to be consistent with change in demand pattern.

In order to analyze the impact of traffic demand on the model performance, three levels of “Degree of Saturation (Volume-to-Capacity ratio)” representing free-flow, normal, and congested traffic conditions are considered as (a) $\frac{v}{c} = 0.35$; (b) $\frac{v}{c} = 0.65$; and (c) $\frac{v}{c} = 0.85$. The intersection delay and total queue length values from optimal signal design are compared with those from Webster’s signal design and the average of three cycles is shown in Table 5.

It can be seen that the average percent delay reduction and queue length reduction are relatively high for “free-flow” traffic condition ($\frac{v}{c} = 0.35$), reflecting better performance of the proposed signal design. In addition, it can be observed that the developed design delivers fairly reasonable performance with average reduction in the range of 5–10% for “normal” and “congested” traffic conditions ($\frac{v}{c} = 0.65$ and $\frac{v}{c} = 0.85$). Therefore, it can be inferred that the proposed signal design yields a better level of service on urban roads relative to the existing design procedures under various traffic conditions.

### Table 4: Percent comparison of Model I relative to Webster’s design across different random seed values.

<table>
<thead>
<tr>
<th>Random seed values</th>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Cycle 3</th>
<th>Cycle 4</th>
<th>Cycle 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R = 12$</td>
<td>12.78</td>
<td>3.17</td>
<td>6.90</td>
<td>12.69</td>
<td>0.56</td>
</tr>
<tr>
<td>$R = 27$</td>
<td>7.65</td>
<td>3.28</td>
<td>14.82</td>
<td>6.67</td>
<td>6.11</td>
</tr>
<tr>
<td>$R = 54$</td>
<td>7.79</td>
<td>6.97</td>
<td>14.99</td>
<td>2.84</td>
<td>3.72</td>
</tr>
<tr>
<td>$R = 66$</td>
<td>10.63</td>
<td>6.60</td>
<td>13.82</td>
<td>12.78</td>
<td>3.13</td>
</tr>
<tr>
<td>Average</td>
<td>9.71</td>
<td>5.00</td>
<td>12.63</td>
<td>8.75</td>
<td>3.38</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>2.46</td>
<td>2.06</td>
<td>3.86</td>
<td>4.87</td>
<td>2.28</td>
</tr>
</tbody>
</table>

### Table 5: Percent comparison of Model I relative to Webster’s design across different degrees of saturation.

<table>
<thead>
<tr>
<th>$v/c$</th>
<th>Cycle</th>
<th>Optimal delay (veh·sec)</th>
<th>Webster’s delay (veh·sec)</th>
<th>Average delay reduction (%)</th>
<th>Optimal queue (m)</th>
<th>Webster’s queue (m)</th>
<th>Average queue reduction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.35</td>
<td>1</td>
<td>4034.3</td>
<td>7164.7</td>
<td>21.85</td>
<td>711.6</td>
<td>1158.5</td>
<td>17.47</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4734.5</td>
<td>5097.6</td>
<td>21.85</td>
<td>832.8</td>
<td>873.0</td>
<td>5.82</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3876.5</td>
<td>4546.2</td>
<td>21.85</td>
<td>676.2</td>
<td>745.1</td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td>1</td>
<td>4881.0</td>
<td>5898.7</td>
<td>6.81</td>
<td>838.3</td>
<td>990.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>7820.1</td>
<td>7964.7</td>
<td>6.81</td>
<td>1314.1</td>
<td>1324.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7524.1</td>
<td>7626.9</td>
<td>6.81</td>
<td>1269.6</td>
<td>1286.1</td>
<td></td>
</tr>
<tr>
<td>0.85</td>
<td>1</td>
<td>9625.9</td>
<td>10263.0</td>
<td>9.50</td>
<td>1622.8</td>
<td>1706.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>10609.3</td>
<td>10808.1</td>
<td>9.50</td>
<td>1782.8</td>
<td>1798.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6516.2</td>
<td>8190.6</td>
<td>9.50</td>
<td>1152.3</td>
<td>1375.7</td>
<td></td>
</tr>
</tbody>
</table>

### 5. Summary and Conclusion

This paper presented a probe-based optimal signal control design that dynamically allocates signal timings based on sample travel time data from Wi-Fi sensors. The proposed traffic control approach eliminates the need to collect data related to macroscopic traffic flow parameters using location-based sensors.

The proposed optimal signal design for an isolated intersection is mathematically modelled based on sample travel time information of probe vehicles using shock wave theory. For the purpose of performance evaluation, an independent formulation is made for the scenario where whole traffic flow information from location-based sensors is available. Results showed that the performance of the probe...
based signal control strategy is comparable to that of the entire stream data being used. The proposed design relative to the existing signal design in the field (i.e., Webster’s design) resulted in a delay reduction of 11.78% when computed theoretically with Webster’s delay equation and 10.41% when implemented in VISSIM. It is also found that the fixed signal timings from Webster’s design led to green wastage, as it is unable to cater to varying demands. On the contrary, it is observed that the proposed formulation effectively accommodates the flow variation and prioritizes high volume approaches in allocating green duration, as is expected from a delay minimization strategy. On evaluation, the performance of the model is found to be consistent under varying patterns of demand and is observed to be benefiting a percent delay reduction in the range of 5% to 20% (across congested to free-flow traffic conditions). Therefore, it can be stated that the proposed model adapts to the fluctuations in demand magnitudes and patterns across cycles, thereby enabling an adaptive signal control using sample travel time data.

It can be concluded that the present study offers a reliable signal control design that renders optimal performance based on real-time variations in demand, addressing the conventional data collection challenges. The novelty of the paper is in the mathematical model and its implementation using sampled Wi-Fi data for signal control. With the ease in data collection and processing procedure in implementing the proposed approach, it can be directly used for real-time traffic signal control design. Extension of this research to relax some of the assumptions made with regard to uniform vehicle arrivals, constant cycle length for all cycles, fixed-phase sequence, and validity for undersaturated and nearsaturated traffic conditions is under progress.

Data Availability

The data used to support the findings of this study are simulation-based and are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors acknowledge the “Connected Intelligent Urban Transportation Lab,” funded by the Ministry of Human Resource Development, Government of India (Project no. CEMHRD008432).

References


