An Adaptive Model for Real-time Estimation of Overflow Queues on Congested Arterials

Liping Fu \(^a\), Bruce Hellinga \(^a\) and Yongliang Zhu \(^b\)

\textbf{Abstract}-- The ability to estimate the status of current traffic congestion of a road network is of significant importance for many Intelligent Transportation Systems (ITS) applications such as in-vehicle route guidance systems (RGS) and advanced traffic management systems (ATMS). Substantial research effort has been dedicated to developing accurate and reliable techniques for estimation of various congestion measures such as link travel time and average travel speed. Few reliable models have however been reported, especially for congested arterials. This paper presents a model that can be used to estimate one of the congestion measures, namely real-time overflow queue at signalized arterial approaches. The model is developed on the basis of the principle of flow conservation, assuming that time-varying traffic arrivals can be obtained from loop detectors located at signalized approaches and signal control information is available on-line. A conventional microscopic simulation model is used to generate data for evaluation of the proposed model. A variety of scenarios representing variation in traffic control, level of traffic congestion and data availability are simulated and analyzed. The evaluation results indicate that the proposed model is promising in terms of the accuracy it can provide and advantages it has over existing models.

\textbf{Index terms}-- traffic, queue, travel time

\section{I. INTRODUCTION}

Quick and reliable estimation of traffic conditions is of critical importance for advanced traveler information and traffic management systems (ATIS/ATMS), of which the common objectives are to provide road users with timely and reliable traffic information and to improve traffic through adaptive tuning of control strategies based on current/predicted congestion. Substantial research effort has been dedicated to developing accurate and reliable techniques for estimation of various congestion measures such as link travel time and average travel speed. Few reliable models have however been developed for signalized arterials, especially under congested traffic conditions. The objective of this research is to develop a real-time model that can be used to estimate one of the congestion measures, namely real-time overflow queue length at signalized arterial approaches.

The problems of estimating queue lengths at signalized intersections have been extensively studied in the literature with the primary objective of developing models for estimating average queue lengths for off-line operational analysis (1-5). Queue estimation has also become a critical component in many traffic signal optimization models such as TRANSYT (6), SCOOT (7) and SCATS (8). In both SCOOT and SCATS, traffic behavior, represented as cyclic flow profiles, are obtained from inductive loop detectors that are located on the approaches of signalized intersections. The flow profiles at individual approaches are then projected to the stoplines of themselves and/or of neighboring intersections and used as arrival flow profiles for estimating and predicting queue lengths. The major disadvantage associated with the SCOOT queuing model is that a saturation flow rate needs to be calibrated carefully otherwise serious errors may result. The SCATS queuing model does not require the knowledge of saturation flow rate as the discharge flow profile is directly available from the loop detector located at the stopline. However, the model requires that the turning ratios at individual intersection approaches be obtained in advance and made available for arrival flow estimation.

Cremer and Henninger (9) presented a model for estimating queue lengths at a signalized intersection approach using information from a loop detector located at an appropriate distance from the intersection. In their model, the approach is divided into small sections and assumed interrelated traffic stream models are established for all sections. Based on the flow conservation law and traffic data at the boundary sections, speed at each section can then be estimated. The estimated speeds are corrected using an Extended Kalman Filter and subsequently used as an indicator of queue status. Although not shown in the paper, their model has the potential advantage of capturing the traffic flow originating inside the link and alleviating the problem that may caused by the queue spilling back to the detector. However, the major disadvantage of their model is that it requires extensive calibration to determine the appropriate underlying traffic model for each specific location. Chang and Su (10) tackled the problem of predicting queue lengths at a signalized approach over a short time horizon. They applied an artificial neural network (ANN) based framework to capture the variation of queue length as related to traffic flow and control variables. After careful training, their ANN models were

\(^a\)Department of Civil Engineering, University of Waterloo, Waterloo, ON, Canada, \texttt{lifu@waterloo.ca}; \(^b\) H. W. Loehner, Seattle, USA
able to provide accurate prediction of the queue lengths for 3 seconds into the future. Similar to the traffic flow model-based approach, this model also requires extensive off-line calibration (or ANN training) for each specific location, and thus has limitation for on-line applications. In this paper, we present a model that aims to overcome this limitation.

This paper is organized as follows. Section 2 outlines the methodologies applied to the development of the model. Section 3 discusses the evaluation of the proposed model under a variety of operating conditions. Finally, conclusions and recommendations are highlighted.

II. THE ESTIMATION MODEL AND ALGORITHM

Consider an intersection approach consisting of a single through lane controlled by a traffic signal. A loop detector is installed at a known distance (D) upstream from the approach stopline to provide point measurements on the traffic flows and speeds at the approach. Due to the cyclic signal interruptions, time-varying queues may form in front of the stopline. In this research, we are interested in estimating the queue length at the time instances when the effective green interval ends and the subsequent red interval starts. This queue length is referred to as overflow queue as it is caused by temporary cyclic overflow resulting from the random fluctuation of arrivals and/or by continuous overflow when the arrival rate exceeds the capacity. The estimation model is intended to provide real-time estimates based on data from the loop detector and the signal control system.

The proposed model is based on the assumption that cumulative arrivals and departures of vehicles at the stopline can be constructed over the time horizon, as illustrated in Figure 1. These cumulative arrival and departure curves are then used to derive the equation for the queue length and delay. Suppose that the estimation process has started at time 0 (the starting time of the first cycle), and is current at cycle i. The overflow queue of the current cycle i, denoted as Q^i, can be obtained through an iterative process on the basis of the conservation of flow, as shown Equation 1.

\[
Q^i = \begin{cases} 
Q^{i-1} + q^i - c^i_a & \text{If } Q^{i-1} + q^{i-1} \geq c^i_a \\
0 & \text{Otherwise} 
\end{cases}
\]  

(1)

where: \( q^i \) = arrival flow during the cycle i at the stopline (pcu) and \( c^i_a \) = lane capacity as determined as the number of vehicles discharged from the stopline during the cycle i under fully saturated conditions (pcu)

For the above model to be operational, the cyclic vehicle arrivals (\( q^i \)) and lane capacity at the stopline (\( c^i_a \)) must first be obtained. The cyclic flow rate arriving at the stopline can be derived from detector data. We propose a scheme to accomplish this, whereby the time-series of vehicle arrivals at the loop detector location are forwarded by an amount of time equal to the free-flow travel time from the detector location to the stopline. The projection is performed using an algorithmic process including the following three steps (see Figure 2):

1. The time series of arrivals at upstream loop detector location (\( A_1 \)), which is aggregated by the detector polling interval (e.g. \( R = 20 \) seconds), is parsed into a series of arrivals with an interval size of one second (\( A_2 \)), assuming that vehicles arrive uniformly within each polling interval.
2. The parsed series \( A_2 \) is projected forward in time by an amount equal to the travel time from the detector location to the stop-line. The travel time can be estimated based on the distance D and a projection speed (V). The result is a new time-series of arrivals at the stop-line (\( A_3 \)).
3. Lastly, the projected arrival series (\( A_3 \)) is aggregated into a series of arrivals according to signal cycles. The estimate for the average arrivals within each cycle (\( q^i \)) can then be determined.

In order to estimate the lane capacity, it is assumed that information on the saturation flow rate, or maximum discharge rate as it is referred to in this paper, and signal timing is available in real time for the given approach. The lane capacity can be estimated using Equation 2.

\[
c^i_a = s \times g^i_e
\]

(2)

where \( g^i_e \) is the effective green interval duration of the cycle i (seconds) and \( s \) is the maximum discharge rate (pcu/second)

The maximum discharge rate is expected to have a critical impact on the performance of this model as any error in this parameter would have a cumulative effect on the final estimates. Consequently, estimation error may accumulate as the process continues moving away from the start point. We therefore introduce an adaptive self-adjustment procedure (SAP) to improve the estimate of the maximum
discharge rate. The SAP is based on the assumption that additional information on whether or not there is a queue present over the detector can be derived from the loop detector data such as speed and occupancy. This information is then used as a basis for increasing or decreasing the maximum discharge rate. For example, if the detector system indicates that the overflow queue has reached the detector, but the queuing model estimates a queue that does not spill back over the detector, then the model has underestimated the queue length, implying a lower maximum flow rate should be used in the model. The structure of the SAP involves the following steps:

1. Estimate the “true” queue reach index (QI) for the current cycle (i) based on the loop detector data (speed and occupancy). If there is a queue over the loop, QI=1; otherwise, QI=0.
2. Estimate the “model” queue reach index (QI*) based on the queue estimate for the current cycle (i) from the queuing model. If the estimate indicates that a queue has reached the loop detector, then QI*=1; otherwise, QI*=0.
3. Compare QI to QI*:
   - if QI ≠ QI*, then
     - if QI > QI* then
       \[ c'_d = c_d + \Delta \text{ for } j=1,2,…i \] and go to Step 4
     - else
       \[ c'_d = c_d - \Delta \text{ for } j=1,2,…i \] and go to Step 4
   - else go to Step 5.
4. Starting from the first cycle, sequentially re-estimate the overflow queue length for cycles \( j (j=1,2,…,i) \) with the updated capacity. Go to Step 2.
5. Stop

We note that instead of adjusting the maximum discharge rate \( (c_d) \), the proposed SAP directly adjusts the cyclic lane capacity \( (c'_d) \) with an increment of \( \Delta \), where \( \Delta \) is a model parameter representing the step size of adjustment.

The above model and the associated algorithm includes three categories of parameters: a) signal timing \( \{c_g, g_e\} \); b) detector location and polling interval \( \{D,R\} \); and c) maximum discharge rate, projection speed and adjustment increment \( \{s,V,\Delta\} \). The first two categories of parameters represent the estimation conditions and can be considered as exogenous factors. The maximum discharge rate and the projection speed are model parameters that need to be selected or calibrated when applied to a given condition. Nevertheless, it can be expected that each of these seven parameters should have certain degree of impact on the performance of the model. The following section attempts to evaluate the sensitivity of the proposed model to some of those parameters, aiming to establish the validity and robustness of the model and the conditions under which the proposed model would be applicable.

III. MODEL EVALUATION

The proposed model has been established on the basis of several important assumptions and includes a number of parameters that need to be provided either in advance (e.g. projection speed, maximum discharge rate) or in real time (e.g. vehicle arrivals, signal timing). The objective of this section is to gauge the estimation quality of the proposed model and its performance sensitivity to the model parameters and variables. The INTEGRATION simulation system (11), which has been extensively validated, was used to generate data under a wide range of operating settings. The simulation exercise includes two networks: one with a single fixed-timed signalized intersection and the other representing an arterial network with three intersections. A sensitivity analysis was conducted using data from the single intersection case to identify critical model parameters, while the three intersection network was used to evaluate the general performance of the model under a realistic traffic environment.

During each simulation run, the following data are recorded for subsequent analysis: time-series of vehicle arrivals at the loop detector location, vehicles passing the stopline and cycle-by-cycle overflow queues representing the observed 'true' queue lengths. The vehicle arrival series produced by the simulation system is then projected downstream to the stopline using the algorithm described in the preceding section. The queuing model is subsequently used to provide the estimates of overflow queue lengths. The quality of the model is evaluated by comparing the time-series of estimated and observed values as well as an aggregated measure, namely the root mean square error (RMS), as defined in Equation 3.

\[
RMS = \sqrt{\frac{\sum_{i=1}^{N} (Q_{\text{observed}} - Q_{\text{estimated}})^2}{N}}
\]

Where \( Q_{\text{estimated}} \) is the estimated overflow queue length at cycle \( i \) (pcu), \( Q_{\text{observed}} \) is the observed overflow queue length at cycle \( i \) (pcu) and \( N \) is the number of cycles

A. Sensitivity Analysis

The simulated network includes a single intersection controlled by a fixed-time signal \( (c_f = 60 \text{ seconds}; g_e = 34 \text{ seconds}) \). A detector with an assumed reliability and accuracy of 100% is modeled on the approach. The vehicle trips are generated with negative exponentially distributed headways at the upstream intersections and no traffic enters or exits at midblock. For each modeling scenario, the network is simulated for 30 minutes starting with zero initial queue. Note that in order to isolate the influencing factors, we used the queueing model without the adaptive adjustment component to estimate the overflow queue.
Determine:

"Model" Queue Reach Index (QI*)

\[ QI = QI^* \]

\[ QI > QI^* \]

Y

N

"True" Queue Reach Index (QI)

\[ c_i^f = c_{i-1}^f + \Delta \]

\[ QI = QI^* \]

N

Y

\[ QI - QI^* \]

N

Y

\[ c_i^f = c_{i-1}^f - \Delta \]

Estimated Overflow Queue, Q^f
1) Sensitivity to Detector Location

The proposed model relies on information from a loop detector to estimate the expected traffic arrivals at the stopline. Therefore, it can be expected that the location of the detector has an impact on the performance of the model. The objective of this section is to analyze the sensitivity of the performance of the model to the detector location. In the simulation, a free-flow speed of 60 km/h and a saturation flow rate of 1800 vehicles per hour are used. Simulation runs were performed for five detector locations (D = 50m, 100, 150m, 250 and 500m) under three levels of congestion (degree of saturation, \( x = 0.95, 1.0 \) and 1.05). The analysis assumes that the detector has a polling interval of two seconds.

The simulated arrival series and the signal times were then used to estimate the queue length. The projection speed was set to 60 km/h and the maximum flow rates were 1810, 1884 and 1910 vehicles per hour for the degrees of saturation of 0.95, 1.0 and 1.05 respectively. Different saturation flow rates were considered to reflect the fact that the actual saturation flow rate used in the Integration model (or equivalent to the maximum discharge rate) is adjusted based on the congestion level with the intention to model the condition-varied discharging rate that are often observed at actual signalized intersection approaches.

Figure 4 shows the relationship between the RMS of the model estimates and detector location under the three levels of demand represented by the degree of saturation. As it can be observed, while the detector location has a noticeable impact on the accuracy of the proposed model, there is a critical point beyond which the impact becomes negligible. For example, when the degree of saturation is less than 1.0, very little additional estimation accuracy was obtained by locating the detector more than 150 meters upstream from the stopline.

![Figure 4](image1.png)

When the detector is located too close to the stopline, significant errors may result under highly congested situation. For example, when \( x = 1.05 \), the estimation error corresponding to the detector location of 150 meters is almost twice as high as the estimation error based on the detector location of 300 meters. This is expected as the higher the traffic demand is and the closer the detector is located to the stopline, the more likely the overflow queue would spill back to the detector. Under this condition, the time series of vehicles passing the loop detector no longer represents the true arrivals at the approach, but rather reflects the capacity of the signalized approach. The implication of this empirical result is that for the purpose of queue estimation, the detector should be located upstream of the farthest reach of the overflow queue. It should be pointed out that, in practice when the detector is located far upstream from the stopline, significant midblock flows may exist between the detector and the signal, resulting in inaccurate estimation of traffic arrivals. The optimal placement of the detector likely involves a trade-off between errors resulting from midblock flows and errors resulting from queue spilling over the detector.

2) Sensitivity to Detector Polling Interval

In this section, we analyze the effect of detector polling interval on the performance of the proposed model. The detector is located 500 meters upstream of the stopline. Simulation runs were performed for the combinations of three polling intervals (\( R = 2s, 10s \) and 20s) and three levels of congestion (\( x = 0.95, 1.0 \) and 1.05). Estimated queue lengths were determined on the basis of the recorded arrival series, a maximum flow rate of 1800 pcu/h and a projection speed of 60 km/h.

Figure 5 illustrates the RMS as a function of the detector polling interval under three degrees of saturation. As it can be observed, there is an expected trend that a shorter polling interval would result in better performance. However, the difference in performance within the range of polling rates from 2 to 20 seconds is practically negligible.
3) Sensitivity to Projection Speed

The proposed model uses the parameter projection speed in estimating the time-series of arrivals at the stopline. It has been proposed that the freeflow speed or posted speed be used when the model is applied to a specific location. However, it is not clear what the magnitude of errors would be when the assumed speed doesn't represent the actual conditions. Similar to the preceding analysis, cases were simulated using Integration under three levels of congestion using a free-flow speed of 60 km/h. Queue lengths were then estimated using projection speeds of 50, 60 and 70 km/h with a maximum discharge rate of 1800 pcu/h.

Figure 6 shows the RMS as functions of the projection speed under three degrees of saturation. It can be observed that, although the performance is insensitive to the projection speed, the free-flow speed of 60 km/h seems to be the optimal choice as it corresponds to the lowest RMS. However, by examining the estimation results at cycle-by-cycle level (Figure 7), the use of freeflow speed did not provide the best estimates in all cycles. This seems to indicate that the actual time taken for vehicles to travel from the loop detector to the stopline varies in response to the size of overflow queue. However, this variation appears to be quite small. It can be concluded that the model is quite robust with respect to the choice of value for the projection speed.

4) Sensitivity to Maximum Discharge Rate

The maximum discharge rate represents the maximum rate at which vehicles can possibly be discharged from the stopline. Because of the sequential nature of the estimation model, use of a maximum discharge rate lower or higher than the actual value may have a compounding effect on the performance of the model. That is, the estimation error may increase as the process moves further away from the starting point. The major objective of this section is to identify the magnitude of such effect. Cases were simulated using a saturation flow rate of 1800 pcu/h. The proposed model is then applied to estimate the overflow queue length using maximum discharge rates ranging from 1600 to 2100 pcu/h.

Figure 8 gives the RMS of the model estimates as a function of the maximum discharge rate. Before drawing any conclusion, it should be pointed out that, the use of a maximum discharge rate of 1800 pcu/h did not yield the lowest RMS under high demand levels ($x>1$). This can be explained as the actual maximum discharge rate used in the Integration simulation model appears to vary somewhat with the level of congestion, thus a maximum discharge rate greater than 1800 pcu/h corresponds to the lowest RMS. It can also be observed that, when the maximum discharge rate used in the queue estimation model is carefully selected, the estimation error can be minimized to an acceptable range. However, the more important evidence shown in Figure 8 is the strikingly high sensitivity of the model performance to the maximum discharge rate. This implies that the model is less robust with respect to this parameter and requires obtaining an accurate estimate of the maximum discharge rate before being applied to each specification. This certainly makes the model less desirable for on-line applications, and also indicates the need for the adaptive, self-correcting scheme described earlier.

5) Effectiveness of the Adaptive Self-Adjustment Procedure

The objective of this section is to gauge the effectiveness of the SAP. Cases were simulated under three levels of traffic congestion ($x = 0.95,1.0, 1.05$) with a saturation flow rate of 1800 pcu/h. The proposed model with SAP is then applied to estimate the overflow queue length using an initial maximum discharge rate of 1600 pcu/h. The detector was located 250 meters upstream from the stopline, and the capacity adjustment step size ($\Delta$) was 0.5 vehicles per cycle.

Figure 9 shows the effectiveness of this adjustment method under a degree of saturation of $1.05$. It is clearly shown that the adjustment scheme has effectively mitigated the problem of error accumulation. Figures 10 and 11 provide the RMS of the model estimates for different traffic demands. It can be observed that, when the SAP is applied, the maximum discharge rate used in the model is automatically adjusted, and the estimation error can be
reduced to an acceptable range. These results indicate that when the SAP is built into the model, obtaining the accurate knowledge of the site specific discharge rate may not be required.

B. Model Performance in a Multiple Intersection Network

The purpose of this section is to examine the performance of the proposed model on a more realistic urban arterial network with multiple signalized intersections. The network consists of three signalized intersections and one un-signalized intersection (stop sign), as shown in Figure 12. Two links were selected for analysis. The first link (Link 1) is the westbound section of Main Street from intersection B to intersection A. It consists of an exclusive left turn lane, an exclusive through lane, and a shared through and right turn lane. A loop detector in the left lane is located at 100 meters from the stop-line. Loop detectors in the through lane and right-through lane are located at 200 meters from the stop line. The traffic signal at intersection A has a cycle length of 90 seconds and a green time interval of 20 seconds for the phase in which WB traffic discharges. The second link (Link 2) is the westbound approach to the intersection of Main Street and Highland Street. The left turn lane loop detector is located 100 meters from the stop-line and loop detectors on the through lane and right-through lane are located at 400 meters from the stop-line. The traffic signal at intersection B has a cycle length of 90 seconds and a green time interval of 40 seconds for the phase in which WB traffic discharges. The simulation is carried out under a set of time-varying traffic demand over a period of 30 minutes. The saturation flow rate and free-flow speed used in simulation are 1800 pcu/h and 60 km/h respectively.

The real-time overflow queue lengths were estimated with and without SAP on the basis of the recorded arrival series, a detector polling rate of 2 seconds, and a projection speed of 60 km/h. The maximum flow rate applied in the queue estimation model were deliberately set very low (1600) for Link 1 and very high (2000) for Link 2.

![Figure 8. Relationship between the RMS of the model estimates and maximum discharge rate](image)

![Figure 9. Comparison of estimated and observed queue lengths: with and without adaptive adjustment (x=1.05)](image)

![Figure 10. Effectiveness of Self-Adjustment Procedure (x = 0.95)](image)

![Figure 11. Effectiveness of Self-Adjustment Procedure (x = 1.05)](image)
Figures 13 and 14 compare the estimated and observed queue lengths at the two intersection approaches, namely Link 1 and Link 2 respectively. It was found that the SAP was invoked at 19th cycle for Link 1, and that one iteration with a capacity increment of 0.5 vehicles/cycle was needed to correct the overestimation problem. Conversely, the underestimation problem for Link 2 was corrected after three iterations of adjustment with a total capacity decrement of 1.5 vehicles/cycle. The average RMS for Link 1 and 2 were 1.8 and 3.9 respectively. Note that this magnitude of estimation errors is expected to be acceptable for the purpose of traffic information provision. The results have further indicated that the SAP is effective in correcting the potential problem caused by the maximum discharge rate and the overflow queue model could be applicable in realistic traffic networks.

This paper has described the development and validation of a queuing model proposed for on-line estimation of queue lengths at signalized intersection approaches. The model was constructed on the basis of the principle of flow conservation, requiring fewer parameters and less calibration effort as compared to some existing models. A self-correcting, adaptive scheme was developed to automatically adjust the maximum flow rate during the estimation process. Simulation experiments were conducted to generate data needed for evaluating the proposed models. The sensitivity analysis has shown that the model is relatively accurate under a wide range of operating environments, and is robust with respect to the polling rate and projection speed. With the integration of the self-correcting scheme, the proposed model was able to accurately track the evolution of real-time overflow queue lengths in realistic traffic environments with any rough estimate of the maximum flow rate. Future research will focus on validating the proposed model using field data and exploring its application in estimating real-time link travel times.

V. ACKNOWLEDGEMENTS
This research was supported by the Natural Sciences and Engineering Research Council of Canada.

VI. REFERENCES