ARTERIAL TRAVEL TIME ESTIMATION FOR ADVANCED TRAVELER INFORMATION SYSTEMS

Wei-Hua Lin, Amit Kulkarni and Pitu Mirchandani

Advanced Traffic and Logistics Algorithm Systems Research Center
Department of Systems and Industrial Engineering
The University of Arizona
Tucson, AZ 85721

(Submitted for Presentation at the 82nd Annual Meeting of Transportation Research Board)

ABSTRACT
While vehicular flows on freeways are often treated as uninterrupted flows, flows on arterials are conceivably much more complicated since vehicles traveling on arterials are not only subject to queuing delay but also to signal delay. Prediction of travel time is potentially more challenging for arterials than for freeways. This paper proposes a simple approach for arterial travel time prediction. The proposed approach decomposes total delay on an arterial into link delay and intersection delay. Intersection delay in the context of arterial travel time prediction is different from the average delay at an intersection. The proposed approach reduces the continuous delay experienced by drivers at each intersection into two distinctive states, a state of zero-delay and a state of nominal delay, coupled with a one-step transition matrix that relates the delay to a through vehicle at an intersection to its delay status at the adjacent upstream intersection. The parameters of the transition matrix are based on three key factors, the flow condition at the intersection, the proportion of net inflows into the arterial from the cross streets, and the signal coordination level. Numerical results show that the model can yield predictions with a reasonable degree of accuracy under various traffic conditions and signal coordination levels.

1. INTRODUCTION
Travel information dissemination is an important component of Intelligent Transportation Systems. Currently, travel information systems usually provide static information between two points calculated solely based on the distance and the speed limit. In reality, travel time varies substantially during the peak and off-peak hours, making the relevancy of static information questionable in many cases. With the rapid improvement in traffic
surveillance systems, vast amount of traffic data are now collected and made available to traffic operations. Prediction of travel time dynamically within a small time window, say, 15 minutes, is now possible. Short-term travel time estimation is useful for travelers to make their route choice decisions, to select different transportation modes, and to determine their departure time from home or from office. The objective of this paper is to develop an arterial travel time prediction model that predicts travel time with a degree of accuracy suitable for these applications and can be adopted by an advanced traveler information system.

While vehicular flows on freeways are often treated as uninterrupted flows, flows on arterials are conceivably much more complicated since vehicles traveling on arterials are subject not only to queuing delays but also to signal delays as well as to delays caused by vehicles entering from the cross streets. Consequently, prediction of travel time is potentially much more challenging for arterials than for freeways. Whereas research on travel time estimation for freeways is very rich (1-12), research on travel time estimation for arterials is quite limited. Sisiopiku et al examined the use of detector output from simulation and field studies to improve the performance of arterial travel time estimation (13, 14). They found that regression equations can be fitted for certain ranges of occupancies to model the observed relationships between arterial through travel time and detector occupancy. However, the method is unable to predict travel time when the queues extend over the detector location.

Travel time on arterials can be decomposed into two components, free flow travel time and delay. Free flow travel time, which is essentially the travel time used for static information, is a function of distance and speed limit. Typically, the average speed tends to slightly exceed the speed limit when traffic is light. For prediction purposes, it is reasonable to use the speed limit or the speed adjusted slightly higher than the speed limit to calculate free flow travel time.

Delay can be further decomposed into link delay and intersection delay. In the case of arterial travel, link delay can arise in two circumstances. First, link delay can be caused by intersection delay. The slow down of a vehicle approaching an intersection could be the result of the driver’s attempt to time its arrival time against the change in traffic signal, in hopes that by the time the vehicle joins the queue the queue will start
moving again. This behavior may trigger a chain reaction in the upstream direction and further slow down the incoming vehicles and temporarily reduce the flow rate on the link. The problem is especially pronounced when a link is relatively short. Clearly, the early deceleration in response to the intersection queue only reduces a vehicle’s waiting time in a standing queue but does not change its overall delay. If we take into account link delay in this case, we may end up double counting the overlap part of link delay and intersection delay. Second, it is perceived by some that link delay may arise spontaneously with the increase in flow. For freeways, link delay is conventionally treated as a function of link flow using the fundamental flow and speed relationship such as the Greenshield function. However, results from recent studies tend to support a near-linear flow and density relationship for uncongested traffic, suggesting that vehicle speeds are near constant before traffic breaks down (15). This seems to be consistent with our own observations of the arterial travel time using cameras installed at two intersections. Our limited data show that link travel time is not very sensitive to link flow when flow remains at medium or high levels. In general, the midblock flow seldom reaches the capacity level because it is metered by upstream intersections. In view of this, we assume that the mid-block link delay is zero. This assumption, however, can be easily removed. We can, for example, distinguish the travel times with only two states, $t^+$ for congested traffic and $t^-$ for uncongested traffic, depending on the flow level. $t^+$ is chosen slightly larger than the travel time based on the speed limit and $t^-$ is slightly smaller. The assumption that the travel time of mid-block link is independent of the flow level reduces our problem from arterial travel time estimation to delay estimation at intersections.

In the past, considerable efforts have been devoted towards characterizing the average and variance of intersection delay. Most of the existing delay formulae assume uniform or Poisson arrivals across the entire cycle. The average intersection delay to through vehicles may not be the same as the average delay at an intersection. The delay an individual vehicle may experience depends on the time the vehicle arrives at the intersection relative to the signal phase. For a poorly coordinated traffic signal system, a vehicle usually arrives at an intersection at any time of the cycle. On the other hand, for a well-coordinated traffic signal system, a green band is pursued to ensure that through
vehicles arrive at the intersection when traffic light is green. Although vehicles are not
guaranteed to be covered by the green band because of platoon dispersion or drivers’
choices of their speeds, they are more likely to arrive at the intersection during the green
phase if traffic signals are operating in a coordinated mode. An arterial travel time
estimation model should take this into account.

For arterial travel time estimation, flow composition could also be an important
factor. Flow composition has little impact on the overall delay under light traffic.
However, when traffic volume is moderately high or high, the impact will grow
substantially. Flows entering the arterial from cross streets may change the service
sequence on the arterial. They may become travel impedance to the arterial flow,
resulting in a higher delay to the arterial through vehicles. For the same level of flow
measured at an intersection, delay to through vehicles traveling on the arterial could be
larger if a significant proportion of flow comes from the cross street, even though the
arterial traffic signals are operating in a coordinated mode. This is illustrated in Fig. 1.
Flows per cycle in both Figs. 1a) and 1b) are identical, so are the signal timing plans in
both cases. In Fig. 1a) every vehicle is delayed. In contrast, none of the vehicles are
delayed in Fig. 1b). However, the average delay at the intersection calculated from both
cases should be exactly the same. In fact, flows that compete with through vehicles do
not have to come from upstream cross streets. They can come from many other sources,
including a mid-block shopping plaza, a parking lot, etc. The arrivals of those vehicles
are not regulated by the signal timing plan and thus independent of the coordination level
of the arterial signal. The presence of entering vehicles from cross streets changes the
arrival distribution, which in turn, changes the average delay to vehicles, especially delay
to through vehicles.

Arterial travel time prediction could adopt a very detailed approach in which one
can dynamically trace the movement of an individual vehicle at every moment and
predict the travel time of that vehicle at each link based on the projected vehicle
trajectory and the predicted vehicle arrival time at each intersection. This approach has
an advantage of potentially yielding an accurate prediction of arterial travel time. It,
however, requires very detailed data, such as the vehicle speed, the queue length at each
approach, and even the acceleration and deceleration profile of vehicles that are either
unavailable or expensive to obtain. Besides, the discrepancy between the actual arrival time and the predicted arrival time of a vehicle at each intersection may propagate in the downstream direction and affect the prediction accuracy for links and intersections further downstream. Given the scarcity of the data sources available on an arterial, we consider the use of more aggregated data, such as traffic volumes at intersections. We also use data summaries that are qualitative in nature (e.g. signal coordination level) but are sufficiently descriptive for the system status. The proposed model is formulated for an arterial that is not fully instrumented to generate real-time data, a case that is true for most of the roads.

The paper is organized into five sections. In the next section, we discuss the formulation of the proposed model. Given the simplification we made for travel time prediction on arterial links, the focus of our model formulation is primarily on characterization of intersection delay. Sec. 3 discusses the property of the proposed model. Closed form solutions are developed for homogenous intersection to exhibit the properties of the proposed model under various conditions. A discussion of the model under heterogeneous intersections is also given. Some numerical examples are presented in Sec. 4. The choice of parameters is discussed. The numerical results show that the model is able to predict the arterial travel time with an acceptable accuracy level. Findings and conclusions are summarized in Sec. 5.

2. THE FORMULATION OF THE PROPOSED MODEL

In order to characterize intersection delay specifically for arterial travel prediction, we consider the following three key parameters obtained from the intersection in question and its upstream intersection:

1) The ratio of the overall flow level to the service capacity for the intersection in question;
2) The net turning movement percentages into the arterial from the cross street at the upstream intersection;
3) The traffic signal coordination level with its upstream intersection.
Qualitatively, travel time increases with the increase in the ratio of the overall flow level to the service capacity. As explained in the previous section, travel time for through vehicles also increases as the net turning movement percentage increases. Travel time decreases as the traffic signal coordination level increases. In the following, these three factors will be explicitly incorporated into the travel time prediction model.

Without loss of generality, we consider first intersections on a one-way arterial with a signal timing plan that has only two phases, a red phase and a green phase. The lost time is incorporated into the red and green by considering the effective red and effective green. The protected left turn phase that prohibits the through vehicle from moving is also treated as part of the red phase. Consider a vehicle driving through a sequence of intersections. Suppose that intersections are labeled with indices \( i = 0, 1, 2, 3, \ldots \), in the direction of vehicle flow. The following is a list of notation to be used in model formulation:

- \( \delta(i) \): an indicator denoting the delay status to the vehicle at intersection \( i \).

\[
\delta(i) = \begin{cases} 
1 & \text{vehicle is delayed at intersection } i \\
0 & \text{vehicle is not delayed at intersection } i 
\end{cases}
\]

- \( d^{(i)} \): predicted delay to the vehicle at intersection \( i \),

- \( E(D^{(i)}) \): nominal delay at intersection \( i \), which is usually computed by a well-defined delay formula.

- \( P^{(i)} \): transition matrix for delay at intersection \( i \).

- \( p^{(i)} \): probability that the vehicle is delayed at intersection \( i \).

- \( p_{\delta(i),\delta(i-1)}^{(i)} \): the conditional probability for a vehicle being at state \( \delta(i) \) at intersection \( i \), given that the vehicle was at state \( \delta(i-1) \) at intersection \( i-1 \).

For an intersection that adopts a fixed timing plan, in the absence of overflow situations, the actual delay a vehicle may experience is a continuous variable, ranging from zero to maximum red. When the overflow situation arises, delay could well exceed the upper bound, depending on the traffic condition. The actual delay depends on the arrival distribution, which is dependent on the signal timing plan of the upstream intersections. Therefore, instead of treating delay as a continuous variable with some commonly assumed probability distribution functions, we consider delay to a through
vehicle with only two states, zero delay or nominal delay. The nominal delay is characterized by a delay formula, which is a function of arrivals, capacity, and saturation flow. In this paper, we use the popular Webster’s formula to represent the nominal delay at every intersection. Our formulation does not require a specific form of delay formula. Other delay formula can also be adopted as well. It is reasonable also to assume that delay at an intersection is conditional on delay at its upstream intersection. With this, we always consider delay to a vehicle at an intersection in conjunction with the delay to the same vehicle at its adjacent upstream intersection. The relationship of delay between the \((i-1)\)th intersection and the \(i\)th intersection can be expressed with the conditional probability defined as follows:

\[
P\{\delta(i) = 1 | \delta(i-1) = 1\} = p_{11}^{(i)}
\]

\[
P\{\delta(i) = 1 | \delta(i-1) = 0\} = p_{10}^{(i)}
\]

\[
P\{\delta(i) = 0 | \delta(i-1) = 1\} = p_{01}^{(i)} = 1 - p_{11}^{(i)}
\]

\[
P\{\delta(i) = 0 | \delta(i-1) = 0\} = p_{00}^{(i)} = 1 - p_{10}^{(i)}.
\]

The determination of the value of \(p\) depends on the three key parameters that will be discussed later. The calculation for delay at each intersection can be conveniently represented by an information tree as shown in Fig. 2. The information tree depicts the probabilistic relationship between the delay status at the current intersection and the delay at all of its downstream intersections. Based on the diagram, one can easily develop the delay prediction formula at each intersection. If a vehicle is delayed at the current intersection, \(i = 0\), then delay at the next intersection is:

\[
d^{(1)} = p_{11}^{(1)} E(D^{(1)})
\]

Delay at the second intersection is:

\[
d^{(2)} = [p_{11}^{(1)} p_{11}^{(2)} + (1 - p_{11}^{(1)}) p_{10}^{(2)}] E(D^{(2)}).
\]

One can continue the process to compute \(d^{(3)}, d^{(4)}, \ldots, d^{(n)}\). Likewise, one can calculate delay for all the downstream intersections if the vehicle is not delayed at \(i = 0\). For homogenous intersections, we can drop the indices that denote specific intersections, i.e., \(p_{11}^{(1)} = p_{11}^{(2)} = p_{11}^{(3)} = \ldots = p_{11}\). By homogeneity, we mean that the intersection pairs have
similar characteristics in terms of v/c ratio, the net flow into the arterial, and the
coordination level of traffic signals. The information tree shown in Fig. 2 can be
conveniently represented by a Markov chain such that delay at intersection \( i \) can be
expressed in the following form:

\[
d^{(i)} = \begin{bmatrix} \delta(i), 1 - \delta(i) \end{bmatrix} \begin{bmatrix} p_{11} & 1 - p_{11} \\ p_{10} & 1 - p_{10} \end{bmatrix} \begin{bmatrix} E(D^{(i)}) \end{bmatrix}.
\]

where \( \begin{bmatrix} p_{11} & 1 - p_{11} \\ p_{10} & 1 - p_{10} \end{bmatrix} \) is a one-step transition matrix. The probability of a vehicle being
delayed at intersection \( i \) is \( p^{(i)} = \begin{bmatrix} \delta(0), 1 - \delta(0) \end{bmatrix} \begin{bmatrix} p_{11} & 1 - p_{11} \\ p_{10} & 1 - p_{10} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \). We can recursively
determine the value of \( p^{(i)} \) for each intersection downstream which result in a set of
dynamic equations:

\[
p^{(1)} = p_{11}, \quad p^{(2)} = (p_{11} - p_{10})p^{(1)} + p_{10}, \quad p^{(3)} = (p_{11} - p_{10})p^{(2)} + p_{10}, \quad \ldots \ldots \quad p^{(i)} = (p_{11} - p_{10})p^{(i-1)} + p_{10}.
\]

The closed-form solution to the conditional delay at intersection \( i \) can be derived
easily based on the above recursive function.

\[
d^{(i)} \bigg|_{\delta(i)=1} = \left[ (p_{11} - p_{10})^{-1} p_{11} + \left( 1 + (p_{11} - p_{10}) + (p_{11} - p_{10})^2 + \cdots + (p_{11} - p_{10})^{i-2} \right)p_{10} \right] E(D),
\]

or equivalently:

\[
d^{(i)} \bigg|_{\delta(i)=1} = \left[ (p_{11} - p_{10})^{-1} p_{11} + \frac{1 - (p_{11} - p_{10})^{i-1}}{1 - (p_{11} - p_{10})} p_{10} \right] E(D)
\]

(1)

If the vehicle is not delayed initially, the conditional delay at intersection \( i \)
becomes:

\[
d^{(i)} \bigg|_{\delta(i)=0} = \left[ \frac{1 - (p_{11} - p_{10})^i}{1 - (p_{11} - p_{10})} p_{10} \right] E(D)
\]

(2)

For a general case in which intersections are non-homogenous, delay at \( i \)th intersection
can be expressed in a matrix form as follows:
The probability of a vehicle being delayed at intersection $i$ is simply

$$p^{(i)} = [\delta(0), 1 - \delta(0)] \begin{bmatrix} p^{(1)}_{11} & 1 - p^{(1)}_{11} & p^{(2)}_{11} & 1 - p^{(2)}_{11} & \cdots & p^{(i)}_{11} & 1 - p^{(i)}_{11} \\ p^{(1)}_{10} & 1 - p^{(1)}_{10} & p^{(2)}_{10} & 1 - p^{(2)}_{10} & \cdots & p^{(i)}_{10} & 1 - p^{(i)}_{10} \end{bmatrix} \begin{bmatrix} E(D^{(i)}) \end{bmatrix}.$$

Equation (3) can be simplified as: $d^{(i)} = p^{(i)} E(D^{(i)})$. The total travel time on a section of an arterial that covers $n$ intersections is $\sum_{i=1}^{n} d^{(i)}$ plus the free flow travel time of that section.

3. PROPERTIES OF THE PROPOSED MODEL

In the following section, we examine the properties of equations (1) and (2) for homogeneous intersections. The purpose of the following discussion is twofold. First, an examination of the properties of the model is useful for model verification, a way to determine if the model makes sense. Secondly, some of the properties can be used as a guideline in selecting the parameters of the model. We will start with considering homogeneous intersections, since the closed-form solution for homogeneous intersections is readily available and the results are transparent. We will then show that the properties examined for homogeneous intersections are also applicable to general situations in which each intersection has a unique transition matrix.

3.1. Homogeneous Intersections

Alternatively, equations (1) and (2) given in the previous section, which was derived for delays at a sequence of homogenous intersections, can be combined into a single expression:

$$d^{(i)} \big|_{\delta^{(0)}=k} = \left[ (p_{11} - p_{10})^{i-1} p_{1k} + \frac{1 - (p_{11} - p_{10})^{i-1}}{1 - (p_{11} - p_{10})} p_{10} \right] E(D^{(i)}), \; k = 0, 1. \quad (4)$$
Note that only the first term depends on $k$, the initial condition. When $i$ is sufficiently large, the first term converges to zero. The second term, however, converges to a constant. It follows that $d^{(i)}|_{\delta(0)=0} \approx d^{(i)}|_{\delta(0)=1}$ for sufficiently large $i$, suggesting that the predicted delay at a remote downstream intersection is independent of the delay status at the initial intersection.

As $i$ increases, the first term vanishes and the second term converges to $\left(\frac{1-p_{11}}{p_{10}} + 1\right)^{-1} E(D^{(i)})$. The range of delay is $0 \leq d^{(i)} \leq E(D^{(i)})$. $d^{(i)}$ approaches 0 if the ratio of $1-p_{11}$ to $p_{10}$ is large. However, $1-p_{11} \gg p_{10}$ is true if both $p_{11}$ and $p_{10}$ are very small, suggesting that the traffic signals for the two intersections are well coordinated.

Ideally, vehicles traveling in a coordinated system should be delayed at most once if the net inflow is negligible. Therefore, $1-p_{11}$, meaning that the conditional probability of a vehicle not delayed at the current intersection given that the vehicle is delayed in the previous intersection, tends to be close to 1. On the other hand, $p_{10}$ is the probability that a vehicle is delayed at the current intersection given that the vehicle is not delayed at the previous intersection, should be much smaller than 1. Consequently, the delay term, $\left(\frac{1-p_{11}}{p_{10}} + 1\right)^{-1} E(D^{(i)})$, should be very small. It follows that the total expected delay, which is the summation of the delay at all intersections, should be small as well.

It is desirable to make $d^{(i)}$ a monotonically decreasing function in $i$. This can be achieved if we choose $p_{11} < p_{10}$. A reversal of this condition would produce a $d^{(i)}$ that oscillates in $i$. Again, this requirement also makes physical sense. If the traffic signal along an arterial is well coordinated, a vehicle that is delayed at the current intersection is better positioned in the vehicle platoon so that the chance for it being delayed again in a downstream intersection is smaller than a vehicle that is not delayed at the current intersection. The rationale behind this is: in a coordinated mode, a vehicle will become better positioned in the green band given that the vehicle is delayed at one intersection. Thus, its chance to be delayed again downstream is smaller.

If traffic signals along the arterial are poorly coordinated, then $1-p_{11}$ tends to decrease and $p_{10}$ tends to increase. In a special case in which traffic signals along the
arterial are completely independent, \( p_{11} = p_{10} = \frac{1}{2} \). The resulting delay predicted for downstream intersections will then increase to \( \frac{1}{2} E(D^{(i)}) \).

The nominal delay used in the model for an intersection, \( E(D^{(i)}) \), is a function of flow, saturation flow level, and traffic signal settings. It is independent of the flow composition. We explained in Sec. 2 the potential influence of flow composition on delay to through vehicles. The impact of a large turning movement percentage on the delay to through vehicles can be represented in the transition matrix at a qualitative level. If the turning movement percentage for \( i-1 \)th intersection is high and the flow level is also high, then both \( p_{10}^{(i)} \) and \( p_{11}^{(i)} \) in the transition matrix should be increased. Consequently, \( d^{(i)} \) should be increased as well. This can be verified with equation (4). However, turning movement percentage should be taken into account in conjunction with the demand condition. At a low flow level, the impact of turning movement on the delay to through vehicles is negligible.

The above discussion shows that the calibration of the transition matrix plays an important role in arterial travel time prediction. Fortunately, the transition matrix only deals with the flow condition, the level of signal coordination, and the amount of turning movement locally, which could reduce the calibration effort substantially.

3.2. Heterogeneous Intersections

Intersections on an arterial may have different characteristics in terms of their flow levels and signal timing plans. An arterial can intersect with another major street or a minor street. In the former case, the resulting intersection is usually called a critical intersection at which flows from all directions are well balanced. In this case, the green times should be equally distributed to all approaches, whereas in the latter case a longer green phase should be given to the arterial than the cross street. Because of the existence of critical intersections, an arterial can be operating in a coordinated mode only for its local sections. In this case, the closed-form solution to delay developed for homogenous intersections no longer apply. However, the proposed model can be conveniently utilized
to handle this situation. For heterogeneous intersections, an arterial can be decomposed into several homogeneous sections based on the similarity in the three key factors. For a sequence of intersections with similar characteristics, they can be treated as homogenous intersections. For intersections that have significantly different characteristics, their transition matrices need to be individually determined. Delays to through vehicles at those intersections, \( d^{(i)} | \delta^{(i)} = k \), can be calculated numerically for each intersection using equation (3).

The effort for a detailed calibration of the transition matrix could be very tedious. For provision of dynamic travel time in an advanced traveler information system, a one-step transition matrix that satisfies the basic properties described in the previous section for homogenous intersections would suffice for the accuracy requirement specific to this type of application. These properties can be used as a set of rule-based guidelines to determine the transition matrix for heterogeneous intersections, obviating the need for a detailed and time-consuming calibration process based on the field data. For example, if intersection \( i \) is well coordinated with its upstream intersection, then the choice of \( p_{11}^{(i)} \) and \( p_{10}^{(i)} \) should be very small. Moreover, \( p_{11}^{(i)} \) should be slightly smaller than \( p_{10}^{(i)} \). If the upstream intersection has a large percentage of the turning flow, then \( p_{11}^{(i)} \) and \( p_{10}^{(i)} \) should be slightly larger than in the situation when the turning percentage is small. Our test results show that minor changes in parameters would not alter the prediction results significantly.

4. A NUMERICAL EXAMPLE

A numerical example is presented here to illustrate the performance of the model. We compare the predicted delay with the simulated delay for an arterial with nine signalized intersections. For simplicity, we consider homogenous intersections with the same signal settings and the same characteristics of arrivals. Arrivals from the cross streets are neglected. The traffic simulation model adopted in this study is the cell transmission model (16). The cell transmission model is a discrete model consistent with the hydrodynamic theory of traffic flow and is especially powerful in capturing the
formation, propagation, and dissipation of queues. The cell transmission model has recently been used for simulating signalized intersection (17,18,19,20).

In the testing example, the component of the nominal delay given in Eq. (1), $E(D^{(i)})$, is represented by the popular Webster delay formula. However, we use only the first two terms of the formula, the deterministic term and the stochastic term, in the following form:

$$\text{Delay} = \frac{1}{2} \frac{R^2}{(1-\rho)C} + \frac{\rho^2}{2(1-\rho)q},$$

where $q$ is the average arrival rate and $\rho$ is a dimensionless parameter, which is often referred to as the degree of saturation. $\rho = 1$ corresponds to the fully saturated condition. $C$ and $R$ represent the cycle length and the red phase, respectively. The one-step transition matrix used for all the intersections has the following form:

$$P^{(i)} = \begin{bmatrix} \frac{R}{C} e^{-\alpha} & 1 - \frac{R}{C} e^{-\alpha} \\ \frac{R}{C} e^{-\beta} & 1 - \frac{R}{C} e^{-\beta} \end{bmatrix}.$$

Four scenarios were simulated, with a combination of different flow and signal coordination levels. The cycle length and the red phase were chosen to be 61 seconds and 27 seconds, respectively. The choice of the parameters $\alpha$, $\beta$, and $\rho$ in each scenario is displayed in Table 1. These parameters were chosen following qualitatively the guideline provided in the previous section. The functional form of the transition matrix given here is kept simple yet satisfies all the properties discussed in the previous section. In practice, the functional form can be fine-tuned according to the characteristics of the relevant parameters and their relationship with delay.

For the four scenarios described in Table 1, ten runs were performed for each scenario. In each run, the travel time of a randomly selected vehicle going through the arterial is recorded. Over ten runs, the travel times of a total of ten selected vehicles are recorded. Fig. 3 plots the trajectory of the selected vehicle in one of the runs. The average travel time is taken and compared with the predicted delay. The results are displayed in Table 2. As shown in the table, the variation of the simulated delay between different
simulation runs for the same scenario is very large. As expected, with all the simplification made for the model, the relative prediction errors are less than 15% in all cases. The prediction is also very robust. In the worst case, the prediction error is less than 20%. Note that the comparison is made for delay, which is only a fraction of the total travel time. The actual prediction errors in terms of total travel time are much smaller.

As shown in Table 2, the difference between the predicted delay and the simulated delay increases as the increase in traffic intensity. The simulation result tends to underestimate the total delay for deterministic arrivals when traffic intensity is very large. Yet the delay formula used in the prediction model tends to overestimate the total delay when traffic intensity is very large. Thus, the discrepancy between the predicted delay and the simulated delay for the high demand case is more pronounced. Even so, the results are still reasonable. However, the examples are not presented for the purpose of model validation. For model validation, field data should be used instead. Actual travel time, perhaps measured from probe vehicles or vehicle re-identification through cameras, should be compared with the predicted travel time.

5. CONCLUSIONS AND RECOMMENDATIONS

The key to predicting arterial travel time is the prediction of intersection delay. Intersection delay in the context of arterial travel time prediction is different from the average delay at an intersection and should be considered in conjunction with some key parameters relevant to the through vehicles on an arterial. A simple approach for arterial travel time prediction was presented in this paper. The proposed approach was developed on the basis of a discrete Markov chain model. By reducing the continuous delay experienced by drivers at intersections into two discrete states, a state of zero-delay and a state of nominal delay, the proposed model formulates the arterial travel time prediction problem into a Markov chain model with a one-step transition matrix. Three key parameters are considered, traffic flow, the signal coordination level between two neighboring intersections, and the flow composition. These parameters have been utilized either qualitatively or quantitatively in determining the conditional probabilities in the
one-step transition matrix in the model. The proposed model exhibits some desirable properties in predicting the arrival time at downstream arterial links. The numerical examples presented in the paper show that the model yields reasonable results for four scenarios with different demand and signal coordination levels. However, the model has some inherent limitations. One of them is that the nominal delay used in the formulation is based on the existing delay formula for intersections. It is well known that many existing delay formula perform poorly for oversaturated intersections. The performance of the model is yet to be improved in this case.

Further validation of the model using the field data, preferably the probe vehicle data, is needed. Besides, the choice of parameters can be studied to streamline the process of calibrating the transition matrix. Work is in progress to extend the model for predicting dynamically the travel time of any two points in a network setting.

ACKNOWLEDGEMENT

The partial financial support of this research by FHWA is greatly appreciated.

REFERENCES


LIST OF TABLES
Table 1: Comparison with model prediction and simulation results
Table 2: Comparison between the simulated delay and predicted delay

LIST OF FIGURES
Figure 1: Comparison of delay to through vehicles with and without turning vehicles
Figure 2: An information-tree representation of the delay status at intersections 1, 2, and 3, given that vehicles are delayed at intersection 0.
Figure 3: Trajectory of a through vehicle in a simulation run (labels on the y-axis are street names).
(a) delay to through vehicles in the presence of vehicles from cross streets (dashed lines: trajectories for vehicles from cross streets; solid lines: trajectories for through vehicles)

(b) delay to through vehicles in the absence of vehicles from cross streets

Figure 1: Comparison of delay to through vehicles with and without turning vehicles.
Figure 2: An information-tree representation of delay status at intersections 1, 2, and 3, given that vehicles are delayed at intersection 0.
Figure 3: Trajectory of a through vehicle in a simulation run (labels on the y-axis are street names).
Table 1: Comparison with model prediction and simulation results

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Traffic Intensity ($\rho$)</th>
<th>$\alpha$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly coordinated, Low demand</td>
<td>0.35</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Highly coordinated, High demand</td>
<td>0.81</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Poorly coordinated, Low demand</td>
<td>0.35</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>Poorly coordinated, High demand</td>
<td>0.81</td>
<td>0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Table 2: Comparison between the simulated delay and predicted delay

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Simulated Delay</th>
<th>Predicted Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly coordinated, Low demand</td>
<td>13.6, 0.0, 17.0, 0.0, 23.8,</td>
<td>7.57</td>
</tr>
<tr>
<td></td>
<td>0.0, 0.0, 0.0, 0.0, 13.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Mean:</strong> 6.80</td>
<td></td>
</tr>
<tr>
<td>Highly coordinated, High demand</td>
<td>17.0, 0.0, 13.6, 10.2, 20.4,</td>
<td>18.38</td>
</tr>
<tr>
<td></td>
<td>10.2, 20.4, 10.2, 20.4, 10.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Mean:</strong> 13.26</td>
<td></td>
</tr>
<tr>
<td>Poorly coordinated, Low demand</td>
<td>78.2, 47.6, 81.6, 44.2, 57.8,</td>
<td>61.07</td>
</tr>
<tr>
<td></td>
<td>44.2, 57.8, 61.2, 57.8, 71.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Mean:</strong> 60.18</td>
<td></td>
</tr>
<tr>
<td>Poorly coordinated, High demand</td>
<td>136.0, 136.0, 142.8, 136.0,</td>
<td>159.43</td>
</tr>
<tr>
<td></td>
<td>139.4, 142.8, 132.6, 132.6,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>132.6, 139.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Mean:</strong> 137.02</td>
<td></td>
</tr>
</tbody>
</table>