RoadAware: Learning Personalized Road Information on Daily Routes with Smartphones

Kang Chen#
Department of Electrical and Computer Engineering
Southern Illinois University, Carbondale, IL, USA 62901
Email: kchen@siu.edu

Haiying Shen
Department of Electrical and Computer Engineering
Clemson University, Clemson, SC, USA 29631
Email: shenh@clemson.edu

***Abstract*—In this paper, we introduce RoadAware, an infrastructure-less system that leverages a smartphone to collect road information on people's daily routes to and from work. Unlike previous research that utilizes moving vehicles equipped with sensors to collect road information, RoadAware does not need the support from infrastructure, road-side facilities or other users and focuses on serving individuals during his/her daily commute to and back from work. RoadAware can provide road information including 1) travel distance and time between traffic lights; 2) duration of red and green signals of each traffic light; and 3) traffic volume on the road between traffic lights. RoadAware builds a model that can deduce the road information based on the wait time and the length of the waiting queue in front of the car when it stops at a traffic light. The road information not only benefits individuals in their daily commutes (e.g., predict waiting at red lights and remind whether there is a delay on daily route) but can also be collected for traffic optimization that otherwise requires costly sensor deployment. We developed RoadAware on Windows smartphones, and our extensive real-world test shows that RoadAware can provide road information with acceptable accuracy.**

***Index Terms*—Road Information, Sensing, Smartphone**

I. INTRODUCTION

Automobile has been a fast and convenient transportation tool in modern society. However, the increasing amount of vehicles also raises many issues and problems on the planning and management of traffic systems. Statistics show that Americans spend over 100 hours a year commuting to work in cars [1]. As a result, there have been active research efforts on large-area road information sensing to improve traffic system efficiency. In these researches, considering it is costly to build and maintain a sufficient number of cameras or sensors on the roadside or at intersections, moving vehicles equipped with various sensors have been utilized as probing nodes to collect traffic information such as traffic conditions (e.g., volume, delay, and fuel consumption) [2]–[6], road status (i.e., potholes) [2], [7], and traffic light status [8], [9].

Though current road sensing systems can effectively and accurately collect general road status, they are not designed specifically for individual drivers. While people are aware of the general situation on their daily commuting routes, they cannot easily acquire accurate data such as traffic volume, traffic light schedules, and travel times and distances between traffic lights. Such data, if obtained, can potentially benefit drivers and the transportation system in several aspects.

First, if people can acquire accurate road information on their daily routes, they can know whether they are delayed on their daily routes and change schedules accordingly. Both applications can reduce driver's frustration caused by slow traffic [10]. Secondly, the durations of red/green light can be used by the Green Light Optimal Speed Advisory (GLOSA) system [11] to provide drivers with speed suggestions to avoid red lights, which reduces fuel consumption and travel time. Thirdly, once network connection becomes available, the road information collected by individual users can be gathered as the input for Intelligent Transportation System (ITS) [12] to improve the overall transportation system efficiency.

Consequently, in this paper, we propose RoadAware, an infrastructure-less and independent system aiming to collect road information on individual drivers' daily commuting routes, which includes (1) travel distance and time between traffic lights; (2) time duration of red and green light signals of each traffic light; and (3) traffic volume on the road between traffic lights. RoadAware is built upon smartphones and only requires the GPS and the clock on smartphones. In recent years, the number of smartphone users has been increasing at a very rapid rate, reaching an 85% year-over-year increase rate. More than 100 million Americans now own a smartphone. Therefore, the popularity of smartphones ensures that the developed system can easily be deployed.

We first model the relationship between the wait time and waiting queue length of a car when it stops at a traffic light, which is related to both the duration of red light signal and traffic volume. Then, RoadAware collects wait time and waiting queue length during daily commuting, identifies datasets obtained with similar traffic volume, feeds them to the model, and deduces the traffic signal duration and traffic volume. RoadAware can be regarded as a self-learning system that learns road information gradually and independently. We implemented RoadAware on Windows smartphones and tested it on daily commutes to campus to show its effectiveness.

The rest of the paper is organized as follows. Section II introduces the design goal and usage condition. Section III and IV introduce baseline techniques and design of RoadAware. The performance of RoadAware is evaluated through real-world test in Section V. Section VI provides the related

#The work was started when at Clemson University.

work. The last section concludes the paper with future work.

II. Preliminary

A. Design Goal

In this paper, we focus on the collection of road information and regard it as the basis for aforementioned potential beneficial applications. Specifically, RoadAware aims to collect the following information on a user's daily commuting route without the support of infrastructure or other users/systems.

- **Information (1)**: The travel distance and average travel time between traffic lights;
- **Information (2)**: The duration of red light signal and green light signal for each traffic light;
- **Information (3)**: The traffic volume on a road between traffic lights, which is measured as the average number of cars passing the traffic light per second.

RoadAware provides Information (1) between any two consecutive traffic lights, including those at which the car turns right or left. For Information (2) and (3), RoadAware can only provide such information for traffic lights where the car passes through in a straight line. This is because whether a car can pass through a left-turn or right-turn traffic light is decided by both the traffic signal and the traffic status on other directions, i.e., a car can possibly turn right with the red light and may not be able to turn left even under the green light. Therefore, additional information is needed to determine the traffic status on other directions, which is out of the scope of this paper.

We exclude the road information during events such as football games and holidays in RoadAware because1) RoadAware aims to provide benchmark data during normal daily commute and 2) people are more sensitive to working day traffic status. As previously mentioned, such measured information can bring about many benefits. With the general traffic information on daily routes, drivers can detect exceptions to better understand current traffic status, which can reduce potential frustration caused by traffic [10]. Also, the collected information can be feed to the Green Light Optimal Speed Advisory (GLOSA) system [11] and Intelligent Transportation System (ITS) [12] for more efficient transportation system.

B. Route Model

We define a **route** R as a set of traffic lights (L) and road segments (S) in sequence:

$$R := \{A_0, S_1, L_1, S_2, L_2, S_2, L_3, \cdots, S_m, L_m, S_{m+1}, E_{m+1}\}$$

where A_0 is the starting point, L_i is the i^{th} traffic light, S_i is the road segment between L_{i-1} and L_i, E_{k+1} is the ending point, and m is the total number of traffic lights. A road segment can contain different types of roads including freeway, highway, and local road. Figure 1 demonstrates a route.

As we only focus on driving experience and cars stop at either a red light or a yellow light, we only consider two light states $\{$*red/stop, green/non-stop*$\}$. For a traffic light, say L_i, the durations of its red and green light signals are denoted by Y_{ir} and Y_{ig}, respectively.

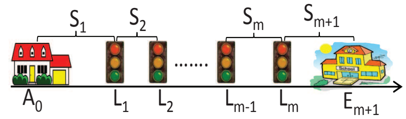


Fig. 1: An example of a home-office route.

There are mainly two types of traffic signal systems:

- *Fixed-time signaling*: the traffic light's signal schedule is pre-determined with fixed red and green light duration.
- *Adaptive signaling*: adaptive signaling control techniques [13] are used to adjust the signal schedule dynamically to improve traffic efficiency. Thus, the durations of red and green light signals change dynamically.

Information (2) in the adaptive signaling system and *Information (3)* on all roads change dynamically at different times. This means it is hard to provide a constant measurement regarding the two kinds of information. In this case, we focus on the average values for each adaptive traffic light on a user's commuting trips. People usually commute to and back from work during approximately the same period of time each day, during which the traffic volume on the road tends to be similar. Then, RoadAware uses the measurement under such an assumption as the **average case** of the road information.

III. Baseline Techniques

A summary of the notations used in this paper is shown in Table I for easy reference.

TABLE I: Notations

R	route to or from work
P_j	j-th sample point on road
L_i	i-th traffic light in route
S_i	road segment between L_i and L_{i+1}
SP_i	start point of S_i
Y_{ir}	duration of red light signal of L_i
Y_{ig}	duration of green light signal of L_i
$\vec{C_i}$	GPS coordinate vector of traffic light L_i
fr	car position measurement frequency
λ_{ia}	rate that cars stop in a waiting queue at L_i
λ_{id}	rate that cars start to move in a waiting queue at L_i
d_{is}	length of waiting queue in front of a car
t_{ia}	time when a car enters a waiting queue at L_i
t_{id}	time when a car starts leaving waiting queue at L_i
t_{irs}	start time of red light signal of L_i where a car stops
T_{irp}	elapsed time of L_i's red light signal when a car stops
T_{iw}	wait time in front of L_i
T_{irgs}	time period a car uses to pass L_i after stopping

A. GPS Reading

To obtain the road information, RoadAware reads the car's positions and speeds from the GPS when the car moves on route periodically, i.e. fr times per second. A larger fr leads to more accurate measurements but also higher energy consumption. We set $fr = 1$ in this paper. As [8], we do not consider energy to be a bottleneck for RoadAware since smartphones can be charged in cars. RoadAware needs to know the GPS positions of traffic lights on the route. We assume such information has already been obtained off-line, e.g., from online map services such as Google Maps [14].

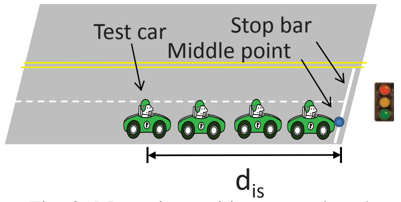


Fig. 2: Measuring waiting queue length.

B. Measuring the Time and Length of a Trip

When a car moves for a period of time, by reading the car's positions from the GPS, RoadAware generates a series of sampling position points $\{P_1, P_2, P_3, \cdots, P_n\}$, where P_j is represented by a GPS coordinator, P_1 is the starting point and P_n is the ending point. Since the timers on smartphones have high accuracy, we directly calculate the travel time by $T = t_n - t_1$, where t_1 and t_n are the time stamps at starting point (P_1) and ending point (P_n) of the trip, respectively.

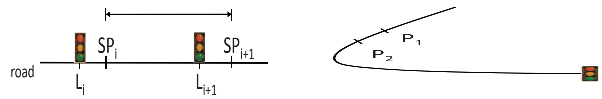
We can measure the length of a trip by accumulating the distances between consecutive GPS positions. However, purely relying on GPS positions may lead to a low accuracy because 1) the GPS may provide insufficiently accurate coordinators [15] and 2) cars are continually moving. Our on road measurement also shows that the error can gets up to 20% on average. To solve this problem, we use the speed to deduce the travel distance, which is more accurate since it considers the Doppler shift in the pseudo range signals from the satellites [16]. Suppose the speed readings at P_j and P_{j+1} are v_j and v_{j+1}, respectively. Then, the distance between the two points is $D(P_j, P_{j+1})$=$(t_{j+1} - t_j)(v_j + v_{j+1})/2$. Finally, the travel distance $L = \sum_{j=1}^{n-1} D(P_j, P_{j+1})$. This method assumes that the car moves steadily or speeds up evenly during $[t_j, t_{j+1}]$, which generally holds since the sampling interval is small (i.e., 1s). The experiment results in Section V-C1 verifies the accuracy of such a method.

C. Detecting Movement Status

Since the GPS can provide driving speed directly, we again use the periodically measured driving speed to detect the movement status. We regard a car as stopped if its speed readings are all less than TH_s meters per second (m/s) for N_s consecutive sampling points, and moving otherwise. Generally, a larger N_s and a smaller TH_s lead to a higher probability of false negative while a smaller N_t and a larger TH_s lead to a higher probability of false positive. After a set of real-world tests, we found that $N_s = 2$ and $TH_s = 2\ m/s$ lead to the best performance with our Windows phones.

D. Measuring the Waiting Queue Length

The waiting queue length is the length of the queue in front of a car when it stops at a traffic light. As shown in Figure 2, each traffic light has a white stop bar in front of it in the moving direction of cars, which is perpendicularly divided into several sections by traffic lanes. Recall RoadAware obtains the GPS coordinates of the middle point of each section with Google Maps off-line. Then, when a car stops in front of traffic light L_i, RoadAware measures the car's distances to these middle points and takes the minimum distance as the waiting queue length of the car, denoted by d_{is}.



(a) Examples of start points of traffic lights.

(b) Two cases of the start point identification.

Fig. 3: Example of possible false positive

IV. The Design of RoadAware

In this section, we introduce how RoadAware collects information to deduce the information listed in Section II.

A. Measuring the Travel Time and Distance

To calculate the travel time and distance of each road segment, RoadAware needs to determine the start point for each segment. As shown in Figure 3(a), RoadAware identifies the sampling point once a vehicle passes traffic light L_i as the start point of road segment S_{i+1}, denoted by SP_i.

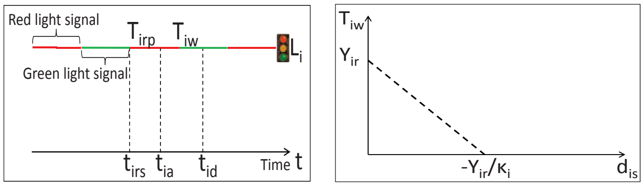
A car's distance to a traffic light usually keeps decreasing before it passes the light and keeps increasing after it passes the light. Accordingly, if two sampling points exhibit increasing distance to traffic light L_i, the latter one is identified as L_i's start point SP_i. However, this method may generate false positives for certain curved roads, as shown in Figure 3(b), in which even when $D(L_i, P_2) > D(L_i, P_1)$, the car has not passed the traffic light. Note $D(L_i, P_j)$ denotes the distance between P_j and L_i, which is measured as the distance between P_j and the middle point of L_i's stop bar.

We add an additional requirement that a SP_i's distance to L_i must be less than a predefined threshold (denoted by TH_p meters) considering the road close to a traffic light usually is straight. In our real-world test, we set TH_p to 80, which leads to a 100% accuracy on start point detection in our road test. Then, RoadAware uses the baseline technique introduced in Section III-B to obtain the travel distance and time between consecutive start points. RoadAware obtains one set of travel time and distance of each route segment each day, and provides the user the average values for reference.

B. Modeling the Waiting Queue at a Traffic Light

When a car stops at a traffic light, we define a car's wait time as the time the car stops in the waiting queue. We define a car's waiting queue length as the length of the queue in front of the car. We model the relationship between the wait time and the waiting queue length, which is related to the duration of red light signal and traffic volume, in this section.

This model assumes that after stopping at the traffic light, the car can leave the traffic light in the next green light. Sections IV-C1 and IV-C2 introduce how to utilize this model to deduce the durations of red/green signals in this case. Section IV-C3 introduces how to deduce such information for traffic lights that do not satisfy such an assumption.



(a) The stop process.

(b) T_{iw} and d_{is}.

Fig. 4: Modeling of the process when a car stops at traffic light L_i.

1) General Model for One Waiting Queue: For a waiting queue at light L_i, we use λ_{ia} and λ_{id} to denote the rate that cars enter and leave the queue respectively. Specifically, λ_{ia} denotes the number of cars that stop at light L_i per second from the moment when it turns red, and λ_{id} denotes the number of cars that leave the waiting queue per second from the moment when L_i turns green. We measured λ_{id} at our local traffic lights and found it is around 0.9 per second. As shown in Figure 4(a), we let t_{irs} be the start time of the red light signal that creates a waiting queue of cars, t_{ia} and t_{id} be the time that a car stops and starts to move in the waiting queue, respectively, T_{irp} be the length of elapsed red light signal when the car stops in the queue, and T_{iw} be the car's wait time in the queue. Thus, $T_{irp} = t_{ia} - t_{irs}$ and $T_{iw} = t_{id} - t_{ia}$.

We also use N_{iw} to denote the number of cars waiting in front of the testing car, denoted car C, in the queue. Then, N_{iw}/λ_{id} is the total time for these cars to start to move when the light turns to green. The remaining duration of the red light when car C stops in the queue equals $Y_{ir} - T_{irp}$. As a result, the wait time of car C at L_i can be calculated as:

$$T_{iw} = Y_{ir} - T_{irp} + N_{iw}/\lambda_{id}. \tag{1}$$

N_{iw} also means the number of cars that stops in the waiting queue before car C during T_{irp}. Therefore, $N_{iw} = T_{irp} * \lambda_{ia}$. Then, Equation (1) is transformed to

$$\begin{aligned} T_{iw} &= Y_{ir} - T_{irp} + T_{irp} * \lambda_{ia}/\lambda_{id} \\ &= Y_{ir} + T_{irp} * (\lambda_{ia}/\lambda_{id} - 1). \end{aligned} \tag{2}$$

On the other hand, N_{iw} can also be deduced by the length of the cars in front of the car: $N_{iw} = d_{is}/\bar{r}$, where d_{is} is the length of the waiting queue in front of the car and $\bar{r}$ denotes average length of a car. d_{is} can be calculated by the baseline technique introduced in Section III-D. Then, we have

$$T_{irp} = N_{iw}/\lambda_{ia} = d_{is}/\lambda_{ia}\bar{r} = d_{is}/\lambda_{ia}\bar{r} \tag{3}$$

Combining Equations (2) and (3), we have

$$T_{iw} = Y_{ir} + d_{is} * \kappa_i \tag{4}$$

where

$$\kappa_i = \mu_i/\beta_i = \frac{\lambda_{ia} - \lambda_{id}}{\lambda_{ia}\lambda_{id}\bar{r}}. \tag{5}$$

Equation (4) describes the relationship between the wait time (T_{iw}) and the length of the waiting queue in front of the car (d_{is}). Since the rate cars start to move usually is faster than the rate cars stop in a waiting queue ($\lambda_{id} > \lambda_{ia}$), κ_i usually is less than 0. Then, T_{iw} has a reserve relationship with d_{is}, as shown in Figure 4(b).

Proposition 1. *Generally, the wait time is reversely related to the waiting queue length; drivers have a short wait time in the queue when the waiting queue is long, and vice versa.*

2) Measuring Model Parameters: The parameters in the model (i.e, Equation (4)) include T_{iw}, d_{is}, λ_{ia}, λ_{id}, $\bar{r}$ and Y_{ir}. RoadAware measures the period of time between car stopping and starting to move as T_{iw}. RoadAware also uses the method introduced in Sections III-D to accurately measure a car's waiting queue length (d_{is}). For one stopping at a traffic light, RoadAware can collect a pair of (T_{iw}, d_{is}) dataset.

Intuitively, λ_{ia} is determined by the traffic volume. In contrast, λ_{id} is not affected by the traffic volume. When a light turns green, the cars in the light's waiting queue leave sequentially. Once the first car sees the green light, it starts to move, and then once the succeeding car sees that its preceding car starts to move, it starts to move. Without loss of generality, we assume that each car takes approximately the same time to move after it sees the green light or the moving of its preceding car. This means that the rates that cars start to move in different queues can be considered as the same. Thus, λ_{id} can be pre-determined. The average length of a car ($\bar{r}$) can also be pre-determined. A study shows that $\bar{r}$ is about 4.12 meters [17]. Considering there are spaces between cars in a queue, we use 4.12+c meters as $\bar{r}$, where c is the length of the estimated average space between stopped cars.

Therefore, only the duration of red light (Y_{ir}) and the rate cars stop in the waiting queue (λ_{ia}) are unknown in the model (Equation (4)). Then, based on Equation (4), given multiple pairs of measured (T_{iw}, d_{is}), we can calculate Y_{ir} and λ_{ia}.

3) A Challenge of Using the Model: We refer to the (T_{iw}, d_{is}) pairs of multiple cars in a given waiting queue at a traffic light as *spatial sample*, and refer to the (T_{iw}, d_{is}) pairs of a given car at a traffic light at different days (queues) as *temporal sample*. Since RoadAware aims to derive road information only from the daily driving records of an individual driver, it can only collect temporal samples.

However, though we assume the traffic status remains stable daily on commuting route, λ_{ia} may vary when the car passes L_i. As a result, given a set of temporal sample (T_{iw}, d_{is}) determined by different λ_{ia}, how to deduce Y_{ir} and λ_{ia}? Below, we introduce a method to deal with this challenge.

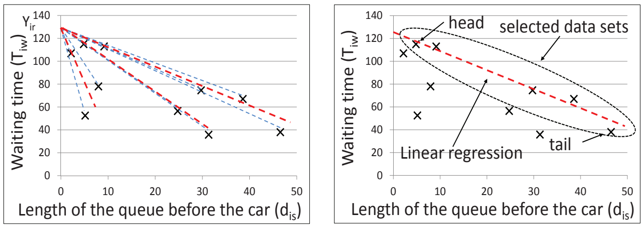
C. Deducing the Duration of Red/Green Light Signals

In this section, we first introduce how to deduce red/green signal durations when the assumption in the proposed model is satisfied in Sections IV-C1 and IV-C2 and then present how to deduce these information when the assumption cannot be satisfied in Section IV-C3.

1) Deducing the Duration of Red Light Signal: If the queue entering rate λ_{ia} remains stable for a given car each day, based on Equation (4), we can derive Y_{ir} and κ_i (determined by λ_{ia}) given two measured datasets (i.e., (d_{is1}, T_{iw1}) and (d_{is2}, T_{iw2})) by solving the equations below:

$$\begin{cases} T_{iw1} = Y_{ir} + d_{is1} * \kappa_i; \\ T_{iw2} = Y_{ir} + d_{is2} * \kappa_i. \end{cases} \tag{6}$$

However, as previously discussed, the queue entering rate λ_{ia} is not the same for a given car each day. The temporal samples collected by RoadAware actually include datasets



(a) Datasets with similar λ_{ia} (b) Deducing the red light duration

Fig. 5: Grouping measured datasets and deducing the length of red light signal.

caused by different λ_{ia} values. To resolve above issue, we examine Equation (4) and Figure 4(b) and find that the waiting queue of a dataset (T_{iw}, d_{is}) is represented by the line connecting point $(0, Y_{ir})$ and (d_{is}, T_{iw}) with slope κ_i, which is decided by the λ_{ia}. This means that the waiting queues experienced by a given car can be represented by lines originating from $(0, Y_{ir})$ with different slopes.

To verify this observation, in our real-world test, we measured Y_{ir} and 10 pairs of (d_{is}, T_{iw}) at a traffic light, which are plotted in Figure 5(a). We connected point $(0, Y_{ir})$ with each (d_{is}, T_{iw}) (blue line) to observe its slope determined by λ_{ia}. We find that based on the slope, the lines (representing waiting queues) can be categorized into three groups (approximation lines in red), i.e., lines in each group have very close slope. This means that queues represented by lines in each group have similar λ_{ia}. Our measured datasets from other traffic lights in the road test also present such a grouping pattern, which is not presented due to page limit.

Therefore, once we can identify one approximation line, we can derive a Y_{ir}. However, we aim to identify all approximation lines and use the average of identified Y_{ir} values as the final result to increase the accuracy. We first introduce how to identify one approximation line to infer Y_{ir} and then discuss how to find all approximation lines.

Approximation line identification: We adopt a heuristic method to find (d_{is}, T_{iw}) datasets with similar car stopping rates (λ_{ia}). It is based on the fact that when the lines connecting $(0, Y_{ir})$ and each of a set of points have similar slopes, these points are generally on the line connecting the most left-up point and the most right-bottom point. As shown in Figure 5(a), each group can also be represented by the line connecting the most top-left point and the most bottom-right point in the group. We can identify this approximation line to infer Y_{ir}. Specifically, among all (d_{is}, T_{iw}) datasets, we call the most top-left dataset as *head* and the most bottom-right dataset as *tail*. That is, the *head* is the dataset that has the longest wait time (T_{iw}) and shortest queue length d_{is}, and the *tail* is the dataset that has the shortest wait time (T_{iw}) and longest queue length d_{is}. RoadAware first identifies the *head* and *tail* from collected (d_{is}, T_{iw}) datasets and then finds all data sets around the line connecting the head and tail.

Dataset (d_{is_j}, T_{iw_j}) is regarded as around the head-tail line if it satisfies $|T_{iw_j} - f(d_{is_j})|/f(d_{is_j}) < TH_c$, where $0 < TH_c < 0.5$ is a threshold for the closeness and $f(\cdot)$ is the function for the head-tail line. Finally, these identified datasets together with the *head* and *tail* form a group of datasets with similar λ_{ia}. Figure 5(b) demonstrates the selected head and tail as well as the identified datasets inside the circle in Figure 5(a).

Linear regression: We then use linear regression to infer Y_{ir}. Suppose the selected datasets include (T_{iw1}, d_{is1}), $(T_{iw2}, d_{is2}), \cdots, (T_{iwn}, d_{isn})$, we calculate κ_i and Y_{ir} by

$$\kappa_i = \frac{\sum_{k=1}^{n}(d_{isk} - \overline{d_{is}})(T_{iwk} - \overline{T_{iw}})}{\sum\limits_{k=1}^{n}(d_{isk} - \overline{d_{is}})^2} \quad (7)$$

and

$$Y_{ir} = \overline{T_{iw}} - \kappa_i \overline{d_{is}} \quad (8)$$

where $\overline{T_{iw}}$ and $\overline{d_{is}}$ represent the average values of T_{iw} and d_{is}, respectively. The dotted red line in Figure 5(b) shows an example of the result of the linear regression.

Recursive approximation line identification: To increase the accuracy, RoadAware continues to identify remaining approximation lines and calculate corresponding Y_{ir} and κ_i pairs. However, lines with fewer than 4 dots are not included in the result since it can easily have a large error. Finally, the average of the all resulted Y_{ir} is considered as the final value. We will see in Section V-B that such a process can improves the deduction accuracy.

2) Deducing the Duration of Green Light Signal: Since cars pass through green traffic lights without stopping, it is difficult to obtain data related to the duration of green signal directly. We propose a heuristic method for this purpose.

When a car stops at a traffic light L_i, the duration of this light's subsequent green light signal should be longer than the time the car uses to pass it. We name this time period as the red-green passing time period, denoted by T_{irgp}. After stopping at a red light, a car usually passes the light after it turns to green and before it turns to red again. That is, $T_{irgp} \leq Y_{ig}$, where Y_{ig} is the green light duration of L_i.

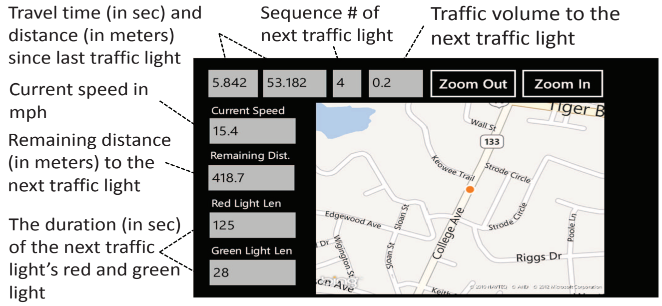
Y_{ig} can also be estimated using another method. It is reasonable to assume that the car arrives at each traffic light randomly. Therefore, the ratio of the number of encountered red signals (denoted N_r) to the number of encountered green signals (denoted N_g) equals the ratio of the length of red signal to that of green signal. Then, for traffic light L_i, $Y_{ig} = Y_{ir} * N_g/N_r$. As a result, given Y_{ir} calculated using the method introduced in Section IV-C1, we can derive Y_{ig}.

RoadAware uses both methods to calculate Y_{ig} and considers the maximum as the final result. That is:

$$Y_{ig} = max\{max\{T_{igrp}\}, Y_{ir} * N_g/N_r\}. \quad (9)$$

3) Traffic Signal Duration at Heavy Intersections: At heavy intersections, a car may not be able to pass the traffic light in one green light, which makes the proposed model infeasible. However, we can obtain the traffic signal duration easier in this case since the car can experience at least one full red light and one full green light in front of the traffic light. In each green light before the car passes through the traffic light, the car moves forward and stops at a new position. The decrease on the waiting queue length during this process can deduce how many cars leave the traffic light during the green

(a) Main interface.



(b) Information panel when the vehicle is on road.

Fig. 6: User interface of RoadAware. *In the main interface (Figure 6(a)), users can click the box showing route name to select route, and press corresponding buttons to view, create, and save road information. When users are ready to go, they just need to press the "Start" button, which would lead the user to the information panel, shown in Figure 6(b). The information panel shows related road information and real-time location (i.e., orange point in the map).*

light, denoted N_{gl}. Then, the green light length is calculated as N_{gl} divided by the car leaving rate, i.e., λ_{id}.

After stopping at the new position, the car has to wait for a full red light plus the time needed for cars in front of it to leave the traffic light when the second green light comes. Then, the red light length can be approximately represented as this period of time minus the time for the cars in front of it to leave the traffic light, i.e., N_{gl}/λ_{id}.

If the car experiences multiple green/red signals, above process repeats for each green/red signal. Then, the final result is the average of these deduced red/green signal durations.

D. Deducing Traffic Volume

For the traffic light that satisfies the model in Section IV-B, we can regard λ_{ia} as the traffic volume. We first simply follow Equation (7) and Equation (5) to calculate λ_{ia} for each identified approximation line. As previously mentioned, λ_{id} is around 0.9 at our local traffic lights. However, each deduced λ_{ia} only represents the car arriving rate at the moment when the car arrives at the traffic light, which may vary in different cases. For example, the traffic volumes on different lanes are slightly different, i.e., left lanes usually are faster. Then, we average the deduced multiple λ_{ia} to represent the average car arriving rate.

For the traffic light that does not satisfy the model in Section IV-B, the car arriving rate is larger than or equal to the car leaving rate λ_{id}. Generally, the larger that car arriving rate, the more time a car needs to pass through the traffic light. Therefore we can regard the $\lambda_{ia} = F_v * \lambda_{id}$, where F_v is a factor to reflect how long a car needs to wait before passing through the traffic light. F_v can be obtained from empirical measurement and is left to future work.

Finally, the overall traffic can be calculated as the average single-lane traffic volume times the number of lanes.

E. Operation Procedure

RoadAware thus collects data through multiple trips. When a car is moving, RoadAware samples a series of positions to measure the travel distance and time between traffic lights (Section IV-A). When a car stops at L_i, RoadAware can detect this stop (Section III-C). It then collects a new dataset (d_{is} and T_{iw}). After it collects sufficient data from multiple trips, it processes the data in the following the procedure in Sections IV-C and IV-D to calculate the durations of red/green signal and traffic volume. Consequently, RoadAware can provide the road information listed in Section II. RoadAware continues collecting data during trips and updating road information.

F. Implementation and User Interface

We implemented RoadAware on Windows phones. Figure 6(a) and Figure 6(b) demonstrate the main interface and the road information panel when the car is on the road, respectively. To use RoadAware, the user first creates their daily routes through the "Create Route" interface. In this process, the user enters the route name and the traffic light sequence number and their GPS coordinate vectors. This information can be obtained off-line. When the user is ready to move, (s)he just needs to choose a route and press "Start". Then, the application switches to the information panel interface in Figure 6(b). On this interface, users are informed of real-time road information, such as the location, speed, traffic volume, travel distance and time since the last traffic light, and remaining distance to the next traffic light. After each trip, the user can view more detailed road information in the "View Route" interface such as the length of each waiting queue d_{is} and the distance and travel time of each route segment.

V. Performance Evaluation

We implemented RoadAware on Windows phones, i.e., HTC Surround and LG QUANTUM. We conducted experiments with a KIA 2005 Optima on three routes, namely *Home2Campus (H2C)*, *Campus2Home (C2H)*, and *LongRoute (LRt)*. The first two routes represent the round trip between campus and one author's home. The third route (LongRoute) is the route from a local gas station to campus. In both figures, the white circles at the ends of the blue routes represent the starting point or ending point of the route, and all other white circles denote the locations of traffic lights. There are 5 traffic lights in routes *Home2Campus* and *Campus2Home*, and 6 traffic lights in route *LongRoute*. In our evaluation, the testing car can pass all traffic lights in the next green signal, which means these traffic lights suit for the model proposed in Section IV-B. The lengths of routes *Home2Campus*, *Campus2Home*, and *LongRoute* are about 1900 meters, 1900 meters, and 4900 meters, respectively, which are obtained through Google map.

These routes represent diverse road conditions and traffic light types. They include busy roads (e.g., the first part of *LongRoute* is on a major local highway) and relatively small roads (e.g., *Home2Campus* and *Campus2Home*). They include traffic lights with fixed signaling (i.e., the third in *Home2Campus* and the fourth in *Campus2Home*) and adaptive signaling (i.e., all other traffic lights). They include traffic lights on major intersections with both directions having high traffic volume (e.g., the second in *Home2Campus* and the fifth in *Campus2Home*) and traffic lights having major traffic in one direction (e.g., the third in *Home2Campus* and the fourth one in *Campus2Home*). Therefore, the experiments show the adaptability of RoadAware on different road conditions.

To simulate the scenario that people do not always commute to or back from work at exactly the same time but roughly the same period of time each day, we ran RoadAware during rush hours 2 to 4 times on each route. Unless otherwise specified, we ran RoadAware on routes *Home2Campus*, *Campus2Home*, and *LongRoute* from 7:30AM to 8:10AM, 4:30PM to 5:10PM, and 5:20PM to 6:00PM, respectively, on each day for 5 consecutive working days. A summary of the experiment settings is shown in Table II.

TABLE II: Experiment settings.

Route	# of traffic light	Length (m)	Test period
H2C	5	1900	7:30AM - 8:10AM
C2H	5	1900	4:30PM - 5:10PM
LRt	6	4900	5:20PM - 6:00PM

We measured the actual durations of red and green light signals of each traffic light with adaptive signaling for several times during the test period each day. We found that they do not change much. Therefore, we use the average of these measurements to check the accuracy of RoadAware. Similarly, the actual traffic volume for each traffic light is calculated as the average number of cars in a waiting queue.

A. Verifying the General Model

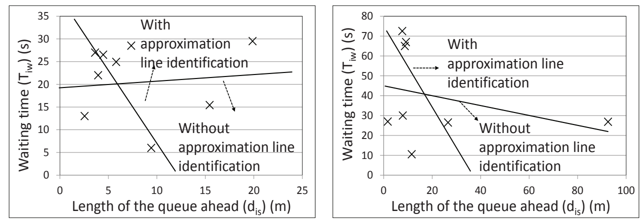
We first validate the model (Equation 4) proposed in Section IV-B1 that depicts the relationship between wait time and distance to a traffic light. For more accurate model validation, we created spatial sample of a waiting queue at a traffic light to deduce Y_{ir} and λ_{ia} based on the model. For each car in a waiting queue, we measured its actual wait time (T_{iw}) and the length of the queue in front of it (d_{is}). Then, we use that data to directly deduce the duration of red light signal and the traffic volume based on Equations (7) and (8).

We selected two traffic lights for test: the second traffic light in route *LongRoute*, named LRt_L_2, and the fourth traffic light in route *Home2Campus*, named $H2C\_L_4$. We measured the data for 5 consecutive red light signals during 9:30AM-10:00AM for LRt_L_2 and during 7:40AM-8:10AM for $H2C\_L_4$, respectively. The deduced data of the two traffic lights are compared with the actual data, as shown in Table III.

We see from the table that the deduced values are very close to their actual values, which demonstrates the correctness of our established model.

TABLE III: Results on verifying the general model.

Traffic light		Red light duration (Y_{ir})	λ_{ia}
LRt_L_2	Deduced	86.7	0.073
	Actual	84.3	0.077
$H2C\_L_4$	Deduced	38.2	0.116
	Actual	40	0.124



(a) L_1 in *Home2Campus*. (b) L_2 in *Campus2Home*.

Fig. 7: Effect of approximation line identification.

B. Verifying Approximation Line Identification

We evaluate the effectiveness of approximation line identification in this section. We compare the results of linear regression with and without the approximation line identification of each traffic light. We found that the regular linear regression without data selection has very poor performance with significant deviations from actual values. Due to page limit, we only present the comparison results for two traffic lights, as shown in Figure 7(a) and Figure 7(b).

We see that 1) there are indeed datasets that have different car stopping rates (λ_{ia}), and 2) when all datasets are fed into the linear regression directly, it leads to inaccurate deduction, as the two labeled lines differ from each other significantly. Such results demonstrate the necessity of approximation line identification and its effectiveness in handling the challenge of using measured parameters in the general model.

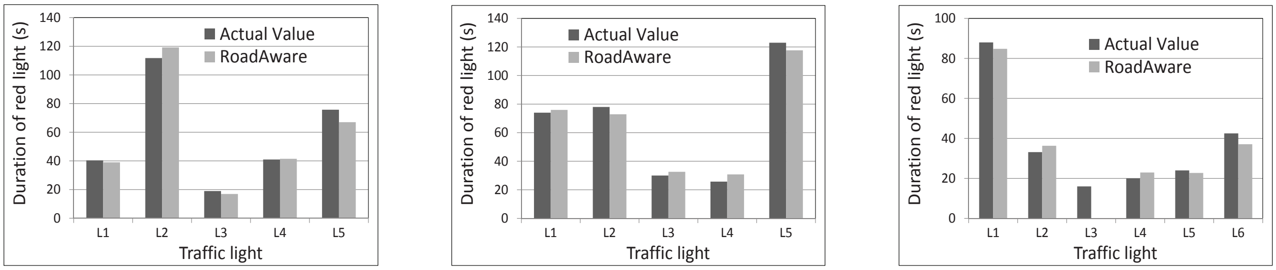
C. Road Information

In this section, we evaluate the correctness and accuracy of the road information provided by RoadAware.

1) Travel Distance and Time: Table IV and Table V show the average of the deduced travel distances and travel times between traffic lights of the three routes. Each row in the table presents the averaged travel distance and time from the previous traffic light to the current traffic light or ending point.

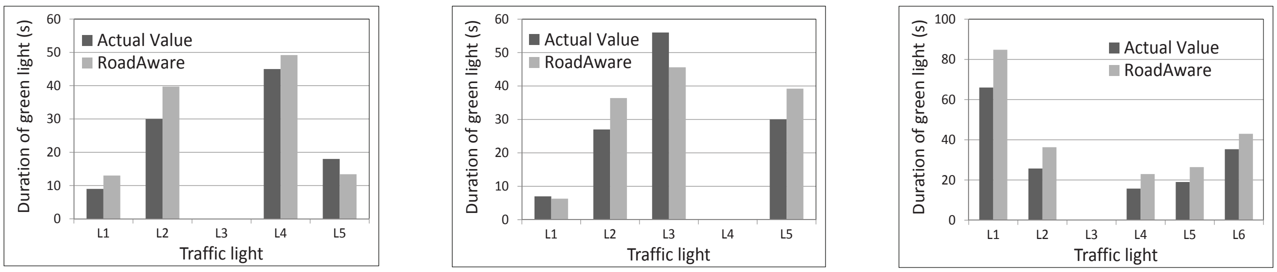
In Table IV, we find that for each route, the sum of the travel distances is very close to the actual value measured by Google Maps. This result demonstrates the accuracy of the baseline method introduced in Section III-B. With such data, users can understand their trips and predict the time needed to travel each route segment, which helps reduce frustration caused by waiting at traffic lights by adjusting people's expectations.

2) The Duration of Red/Green Light: Figures 8(a), 8(b), and 8(c) present the deduced value and actual duration of red light of each traffic light in routes *Home2Campus*, *Campus2Home*, and *LongRoute*, respectively. We see from the three figures that the deduced red light durations for almost all traffic lights are very close to their actual values. In *LongRoute*, RoadAware failed to deduce L_3's red light duration since L_3 turns red only occasionally. These experimental results further verify the correctness of the model proposed in Section IV-B1



(a) Home2Campus. (b) Campus2Home. (c) LongRoute.

Fig. 8: The duration of red lights on the three routes.



(a) Home2Campus. (b) Campus2Home. (c) LongRoute.

Fig. 9: The duration of green lights on the three routes.

TABLE IV: Travel distance of each route segment (m).

Traffic light	H2C	C2H	LRt
L_1	181.6	106.7	767.8
L_2	226.6	245.4	600.6
L_3	523.6	270.3	1299.7
L_4	325	311.2	726.3
L_5	271.5	521.2	897.3
L_6	350.4	416.6	436.3
L_7	-	-	195.7
Total	1878.7	1871.4	4923.7
Actual Total	1900	1900	4900

TABLE V: Travel time of each route segment (s).

Traffic light	H2C	C2H	LRt
L_1	29.1	66.5	110.1
L_2	49.9	72.7	66.2
L_3	39.8	37.1	89.9
L_4	39.3	24.3	50.7
L_5	91.1	67.9	67.9
L_6	28.7	35.8	42.5
L_7	-	-	23.5
Total	277.9	304.3	450.8

and the effectiveness of RoadAware in providing red light duration data on individuals' daily routes.

Figures 9(a), 9(b), and 9(c) present the deduced green light durations of each traffic light in route *Home2-Campus*, *Campus2Home*, and *LongRoute*, respectively. We see that the accuracy of the deduced green light duration lies in the range of [70%, 90%]. This is because we only use a heuristic model to deduce the duration of green light. Though not as high as that of the deduced red light duration, such an accuracy rate is satisfying since it reflects the rough durations. Additionally, drivers are usually less sensitive to the duration of green lights than to the duration of red lights, as green lights require no waiting. Therefore, the deduced green light duration with slightly lower accuracy is still helpful to them.

We also see that RoadAware failed to deduce the green light duration for L_3 in route *Home2Campus*, L_4 in route *Campus2Home*, and L_3 in route *LongRoute*. These failures are caused by the same reason as explained previously.

3) Traffic Volume: Tables VI, VII, and VIII compare the traffic volumes at each traffic light deduced by RoadAware to their actual values measured by us. We used the rate that cars stop at the waiting queue as the traffic volume in RoadAware. Then, for the purpose of comparison, we calculated the measured traffic volume as the average number of cars stopped at a red light divided by the duration of the red light. We see from the three tables that RoadAware can accurately deduce the traffic volumes at most traffic lights. Even with some deviations (i.e., L_4 in route *Home2Campus* and L_1 and L_2 in route *LongRoute*), the deduced traffic volumes still reflect the degrees of traffic volumes in different road segments.

TABLE VI: Traffic volume at each L_i in *Home2Campus*.

	L_1	L_2	L_3	L_4	L_5
RoadAware	0.090	0.067	0.131	0.234	0.130
Actual Value	0.102	0.072	0.140	0.232	0.119

TABLE VII: Traffic volume at each L_i in *Campus2Home*.

	L_1	L_2	L_3	L_4	L_5
RoadAware	0.068	0.218	0.243	0.244	0.116
Actual Value	0.059	0.245	0.233	0.252	0.103

TABLE VIII: Traffic volume at each L_i in *LongRoute*.

	L_1	L_2	L_3	L_4	L_5	L_6
RoadAware	0.22	0.22	-	0.12	0.14	0.12
Actual Value	0.25	0.23	0.13	0.13	0.13	0.15

4) Summary: In summary, this one-week experiment (around 13 runs) on three test routes demonstrates that Road-Aware is able to successfully deduce three sets of road information indicated in Section II. While the accuracy of green light duration and traffic volume is slightly low, the trends and differences among different traffic lights or road segments are clearly reflected. These results prove the ability of RoadAware in an individual's smartphone to effectively deduce information on the user's daily route without requiring infrastructure or collaboration from other users.

VI. Related Work

Recent research has proposed using moving vehicles equipped with sensors (including smartphones) as probing nodes for various sensing applications [2]–[7], [9], [18]–[20]. Hoh *et al.* [18] used virtual trip lines to preserve the privacy of probing vehicles. VTrack [3] combines WiFi signals and the GPS data to save energy during delay sensing. Nericell [2] utilizes multiple sensors on smartphones (i.e., accelerometer, microphone, GSM radio, and GPS) to sense rich on-road information such as potholes, bumps, braking, and honking. SignalGuru [8] is a GLOSA system. It utilizes cameras to detect the transition of traffic signals and disseminates the information to other vehicles in order to provide speed suggestions and avoid red lights. POVA uses reported vehicle positions and speeds to decide traffic light status (i.e., red or green) in urban areas. SEER [4] uses taxi sensory data sensed by the GPS, which tends to be noisy and lossy, to infer the traffic conditions at any site in Shanghai. The work in [5] also utilizes probing vehicles for urban traffic sensing. It uses compressive sensing to infer hidden structures in sensory traffic data and solves the problem of data vacancies caused by limited number of probing vehicles and uneven distributions of cars. In ParkNet [19], the authors deployed a GPS and an ultrasonic rangefinder on a vehicle to detect available roadside parking spaces. Eriksson *et al.* [7] utilized vibration and GPS sensors to detect potholes and other severe road surface anomalies though probing vehicles. GreenGPS [6] uses fuel consumption sensors to detect the fuel consumption status of cars on different roads, which is further utilized to guide drivers and reduce fuel consumption. CityDrive [20] uses information collected from vehicles to deduce red/green light length and provide speed suggestions to drivers.

RoadAware is motivated by several previous works [8], [9]. However, unlike these applications, RoadAware aims to collect road information for individual drivers during their daily commutes in an infrastructure-less and independent manner.

VII. Conclusions and Future Work

In this paper, we propose RoadAware, an infrastructure-less and independent system built on a user's smartphone to provide personalized road information on his/her daily route to and from work. RoadAware uses the GPS and clock on the smartphone to record the information on position and time when the car is moving or stopped at traffic lights. Based on such information, RoadAware can deduce the durations of red and green light signals and traffic volumes on the road using our established model. We implemented RoadAware on Windows phones. Extensive real-world experiments demonstrate the effectiveness and accuracy of RoadAware. Also, the core function of RoadAware only requires a GPS and a clock, enabling RoadAware to be easily ported to other GPS-enabled devices and adopted by the general public.

RoadAware can help individual drivers to better understand their daily routes and adjust their trip expectations. It can also benefit the GLOSA system, distributed traffic monitoring and traffic signal optimization. These promising potential applications will be the focus of our future work.

Acknowledgements

This research was supported in part by U.S. NSF grants NSF-1404981, IIS-1354123, CNS-1254006, IBM Faculty Award 5501145, Microsoft Research Faculty Fellowship 8300751, and the startup fund from SIU.

References

[1] "Average car commuting time." http://usgovinfo.about.com/od/ censusandstatistics/a/commutetimes.htm.

[2] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: using mobile smartphones for rich monitoring of road and traffic conditions." in *Proc. of SenSys*, 2008.

[3] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson, "VTrack: accurate, energy-aware road traffic delay estimation using mobile phones." in *Proc. of SenSys*, 2009.

[4] H. Zhu, Y. Zhu, M. Li, and L. M. Ni, "SEER: Metropolitan-scale traffic perception based on lossy sensory data." in *Proc. of INFOCOM*, 2009.

[5] Z. Li, Y. Zhu, H. Zhu, and M. Li, "Compressive sensing approach to urban traffic sensing." in *Proc. of ICDCS*, 2011.

[6] R. K. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. F. Abdelzaher, "GreenGPS: a participatory sensing fuel-efficient maps application." in *Proc. of MobiSys*, 2010.

[7] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The pothole patrol: using a mobile sensor network for road surface monitoring." in *Proc. of MobiSys*, 2008.

[8] E. Koukoumidis, L.-S. Peh, and M. Martonosi, "SignalGuru: leveraging mobile phones for collaborative traffic signal schedule advisory." in *Proc. of MobiSys*, 2011.

[9] X. Liu, Y. Zhu, M. Li, and Q. Zhang, "POVA: Traffic light sensing with probe vehicles." in *Proc. of INFOCOM*, 2012.

[10] P. Koonce, L. Rodegerdts, L. Kevin, S. Quayle, S. Beaird, C. Braud, J. Bonneson, P. Tarnoff, and T. Urbanik, "Traffic signal timing manual." in *U.S. Department of Transportation*, 2008.

[11] K. Katsaros, R. Kernchen, M. Dianati, and D. Rieck, "Performance study of a Green Light Optimized Speed Advisory (GLOSA) application using an integrated cooperative ITS simulation platform." in *Proc. of IWCMC*, 2011.

[12] "The intelligent transportation systems for traffic signal control deployment benefits and lessons learned." http://ntl.bts.gov/lib/jpodocs /brochure/14321_files/a1019-tsc_digital_n3.pdf.

[13] "Adaptive signal control techniques." http://www.fhwa.dot.gov/every daycounts/ technology/adsc/.

[14] "Google map." http://maps.google.com/.

[15] "Gps accuracy." http://www.gps.gov/systems/gps/performance/accuracy/.

[16] "Gps based speedometer." http://gpssystems.net/accurate-gpsbased-speedometer/.

[17] "Average car length." http://wiki.answers.com/Q/What_is_the_average_length_of_a_car.

[18] B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J. Herrera, and A. Bayen, "Virtual trip lines for distributed privacy-preserving traffic monitoring," in *Proc. of MobiSys*, 2008.

[19] S. Mathur, T. Jin, N. Kasturirangan, J. Chandrasekaran, W. Xue, M. Gruteser, and W. Trappe, "ParkNet: drive-by sensing of road-side parking statistics." in *Proc. of MobiSys*, 2010.

[20] Y. Zhao, Y. Zhang, T. Yu, T. Liu, X. Wang, X. Tian, and X. Liu, "Citydrive: A map-generating and speed-optimizing driving system," in *INFOCOM, 2014 Proceedings IEEE*. IEEE, 2014, pp. 1986–1994.