Determining Time to Traverse Road Sections based on Mapping Discrete GPS Vehicle Data to Continuous Flows

Jeffrey Miller[1], Sun-il Kim[1], Muhammad Ali[2], Timothy Menard[1]
[1]Department of Computer Systems Engineering, [2]Department of Mechanical Engineering
University of Alaska, Anchorage
{jmiller, kim, ali, astsm12}@uaa.alaska.edu

Abstract – **In this paper, we present and analyze an algorithm for mapping discrete GPS data gathered from vehicles to a continuous flow of data to determine the time to traverse a road section. Vehicle-tracking devices are installed in 80 probe vehicles in the Anchorage area, and a specific roadway section was chosen as a test section. Drivers for this study drove from before the start of the test roadway section past the end of the test roadway section, measuring the time to travel from the start to the finish of the test roadway section. The vehicle-tracking devices report speed and location every 10 seconds. From this data, we calculated the amount of time to traverse the test roadway section using our proportional model and compared it to the actual amount of time it took to traverse the test roadway section. We performed the analysis assuming the vehicle-tracking devices were reporting location every 10 seconds, 20 seconds, 30 seconds, 40 seconds, 50 seconds, and 60 seconds. With an average actual time to traverse the test roadway section of 2 minutes 28 seconds, the error rate based on the proportional model was between 1.8%-9.2% (2.7-13.1 seconds), based on how frequently the vehicle was reporting its location. Merely taking the average speed on the edge from the vehicle reporting its speed and location during those same durations had an error rate between 14.2%-25.8% (24.7-41.1 seconds). Our results show that the proportional model has a small error rate (1.8% with 10 second reporting time) and can accurately represent the time to traverse roadway sections based on discrete readings from a small number of probe vehicles.**

I. INTRODUCTION

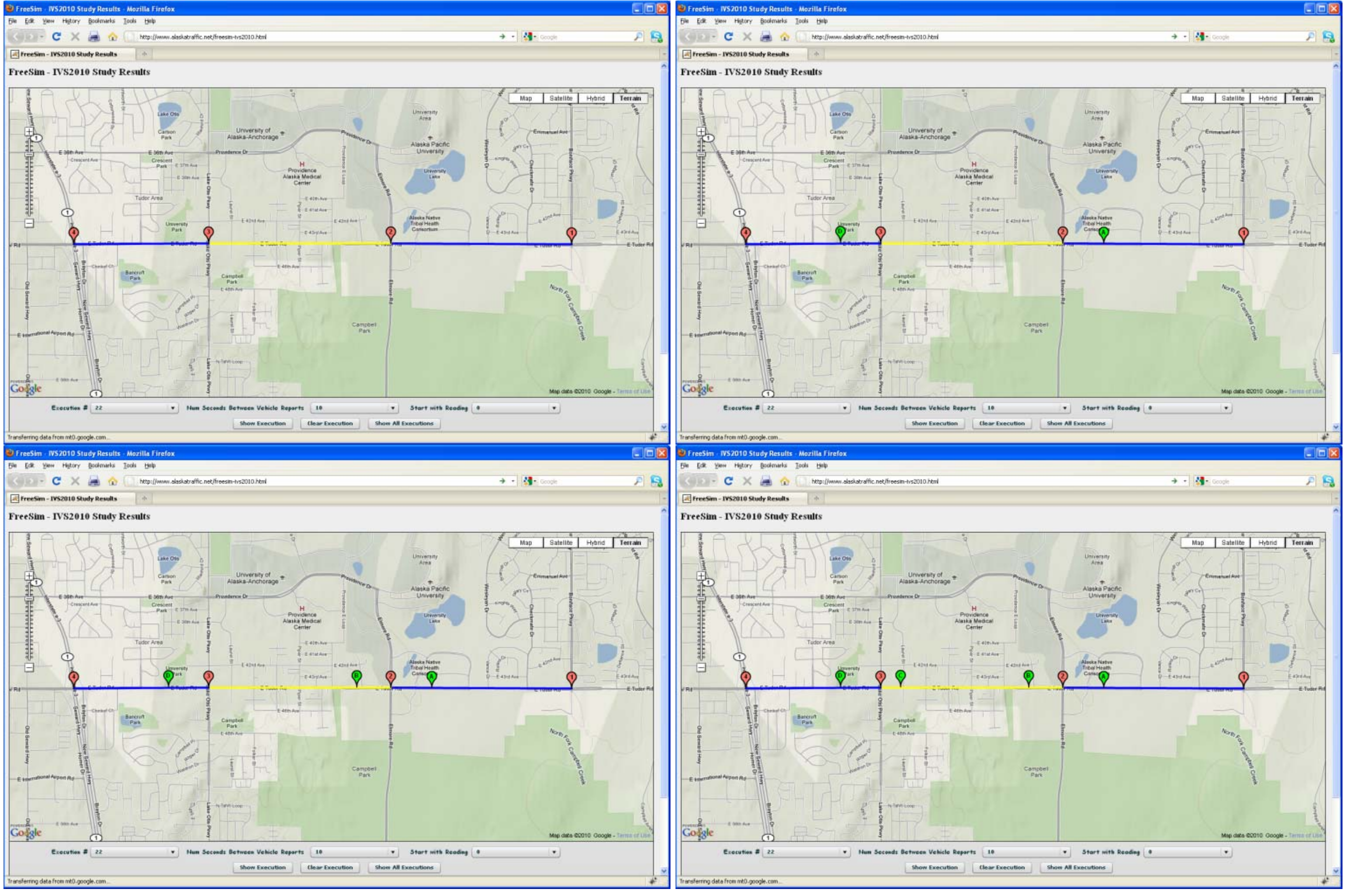
Gathering the speed and location of individual vehicles has become essential for many Intelligent Transportation System (ITS) applications that have been proposed. Determining the time to travel from one location to another has been attempted using many discrete forms of technology, including inductor loops, video cameras, and speed sensors, though no discrete technology can provide as much information about a vehicle as a device that is traveling with the vehicle. Probing devices, such as cellular phones, navigation systems, tracking devices, and mobile communication devices, provide a means for communicating with other vehicles or some roadway infrastructure from within a moving vehicle. The architectures employed for this communication are typically vehicle-to-infrastructure (V2I, sometimes referred to as vehicle-to-roadway, V2R) or vehicle-to-vehicle (V2V). With a V2I architecture, the vehicles will communicate with some central computing device that is connected through a roadway infrastructure, such as the cellular network. The data that vehicles transmit have typically included speed, location, and direction. As additional data is able to be retrieved from vehicles, new applications are being developed, though the network through which the data is gathered is still based on a V2I architecture.

Many applications of ITS data have been proposed, including fastest path determination, incident identification, origin-destination (OD) matrix determination, emergency vehicle routing, signal timing modification, ramp metering, and congestion relief, among others too numerous to list. The common theme among all of these proposed applications is the lack of data available from vehicles. Obtaining static data from discrete devices has been an acceptable substitute for obtaining data from individual vehicles in real-time, but technology has advanced to the point of being able to gather data from vehicles in a distributed manner. The current applications can now be improved upon based on the availability of distributed vehicle data.

In Anchorage, there are currently 80 vehicles equipped with tracking devices that report the speed, location, and direction of the vehicle to a central server every 10 to 60 seconds. Based on this data, a map is generated that provides drivers with the amount of time to traverse arterial roadways in Anchorage. The current means of determining the time to traverse a roadway is based on dividing the distance of a roadway by the average speed of a vehicle traveling along that roadway, known as the speed model. Since the majority of roads in Anchorage are traffic-regulated (as opposed to free-flowing), vehicles report a speed of 0 when stopped at a traffic signal. This obviously has a huge impact on the average speed along that roadway, and typically provides a time to traverse that is substantially different from the actual amount of time.

In this paper, we propose an algorithm that is based on location and time rather than speed. Since the vehicles are reporting their speed *and* location every 10 seconds, our algorithms uses a proportional model to map the time to traverse from one reported location to another to an edge in our transportation graph. Taking a single edge in our graph as a sample, drivers of the probe vehicles measured the actual amount of time to traverse the sample roadway represented by the chosen edge. The actual amount of time was then compared to the amount of time calculated using the proportional model and the speed model. Assuming some

FIGURE I. FREESIM SCREENSHOTS SHOWING (FROM TOP-LEFT CLOCKWISE) a) TEST ROADWAY SECTION USED, INCLUDING PREVIOUS AND NEXT EDGES b) 0 VEHICLE TRANSMISSIONS ON EDGE FROM 2 TO 3, c) 1 VEHICLE TRANSMISSION ON EDGE FROM 2 TO 3, d) 2 VEHICLE TRANSMISSION ON EDGE FROM 2 TO 3



FreeSim - IVS2010 Study Results

Execution # 22

Num Seconds Between Vehicle Reports 10

Start with Reading 0

vehicles are not transmitting data every 10 seconds, we also analyzed this data by increasing the duration between vehicle reportings to 20 seconds, 30 seconds, 40 seconds, 50 seconds, and 60 seconds. The results from these studies are presented in this paper as follows. Section II describes related work for gathering vehicular data and determining the time to traverse sections of a roadway based on ITS data. Section III explains the configuration of vehicles and the roadway used in this study. Section IV provides an overview of the speed model and proportional model algorithms for determining the time to traverse a roadway based on distributed vehicle data. Section V analyzes the results obtained from the two models, and the conclusion is in section VI.

II. RELATED WORK

With the current number of loop detectors and video cameras installed along U.S. highways, discrete identification devices have become the most used technology in detecting traffic patterns. Studies have concluded that the more space there is between loop detectors, the more inaccurate the reported congestion level [6]. Even with advanced algorithms, loop detectors can still produce extreme errors. Using single or double loop estimations can produce inaccuracies of 25-30% [5]. However, using an adaptive algorithm to cope with the fluctuation of congestion significantly reduces the error percentage [8]. Nevertheless, even with these advances in loop detector algorithms, it has been shown through checking the collected information against probe vehicle data that during times of congestion either from traffic, incidents, or construction, loop detectors produce data with a much greater percentage of error [4]. Further, closed circuit television (CCTV) cameras along with loop detectors have estimated traffic speed with an error percentage around 10% [7].

With loop detectors being prone to errors and with the advancement of global positioning systems (GPS), it is now easier to track vehicles in continuous intervals and collect their speed. This allows for a map to be constructed that represents real-time traffic. Using probe vehicles, map matching has been combined with an algorithm for computing speed from point-to-point locations to allow accurate mapping of the traffic [3]. Probe data by itself during peak traffic times can cause prediction errors to increase due to more probes in close proximity to each other [1]. In Shanghai, researchers have combined probe vehicle and loop detector data to produce a very precise map with an accuracy of 97.5% [12].

While vehicle GPS sensors are not as common as cell phones, there have been major developments in using the GPS functionality of smart phones to determine real-time traffic conditions. Florida International University found through their investigations that using cell phones as vehicle probes is feasible under normal traffic flow conditions for travel time estimation. However, their research also found that the cell phone probing is not as accurate in congested traffic conditions, and the accuracy decreases rapidly as the congestion increases [9]. The authors of [10] attempted to prove that through the use of smart phones that accurate speed and travel times could be produced in real time. Waze, a real-time traffic application for smart phones, is currently testing the GPS capabilities of cellular phones to report live traffic reports to users based on speed calculation algorithms [11]. Studies from some projects, including MIT's CarTel [13] and UC Berkeley's Mobile Millennium [14] projects, have begun publishing results and providing access to the data gathered through cellular phones used as vehicle probes.

Through all of the studies that have been conducted using probe vehicles, the accuracy of determining the time to traverse roadway sections has fallen short. The results obtained from our proportional algorithm provide for a more accurate map of real-time traffic conditions to be created.

III. ENVIRONMENT

The transportation network in Anchorage has been represented as a weighted directed graph with nodes identified as intersections and edges as the roadways connecting the intersections. The weight on each edge represents the time to traverse the roadways represented by the edge based on the current flow of traffic. As the speed on a roadway can change constantly, the algorithm used for determining the weight on the edge must be dynamic enough to withstand nearly-constant updates. Further, as the number of edges only for main arterials in Anchorage (with a speed limit of greater than 35mph/56kph) is over 1000, the algorithm must be efficient and not consume a substantial amount of resources, including CPU cycles.

In Anchorage, there are 80 vehicles equipped with tracking devices that report speed, location, and direction every 10 to 60 seconds. This data is transmitted through a V2I architecture over the cellular network to a central server hosted at the University of Alaska, Anchorage. After the server receives this data, it stores it in a database for historical lookups. The real-time data is also fed directly into an algorithm for determining the amount of time to traverse a roadway (or compute the weight of an edge). The two algorithms we have compared in this study are the speed model and proportional model, described in the next section.

We chose to compare these algorithms on Tudor Road between Elmore Road and Lake Otis Parkway, a 0.99 mile/1.59 kilometer roadway section in Anchorage that becomes heavily congested during the rush hour times of the day (typically between 7:00a.m.-9:00a.m. and 4:30p.m.-6:30p.m.). The roadway section is shown in Figure I.a, identified by the edge between nodes 2 and 3. The edge before (between nodes 1 and 2) and after (between nodes 3 and 4) the test roadway section are necessary for computations in the proportional algorithm. The FreeSim application [2,15], which uses Google Maps [16] was used to test the algorithms and analyze the results, and the raw data used for this study can be visually seen and obtained from http://www.alaskatraffic.net/freesim-ivs2010.html.

IV. SPEED MODEL AND PROPORTIONAL MODEL ALGORITHMS

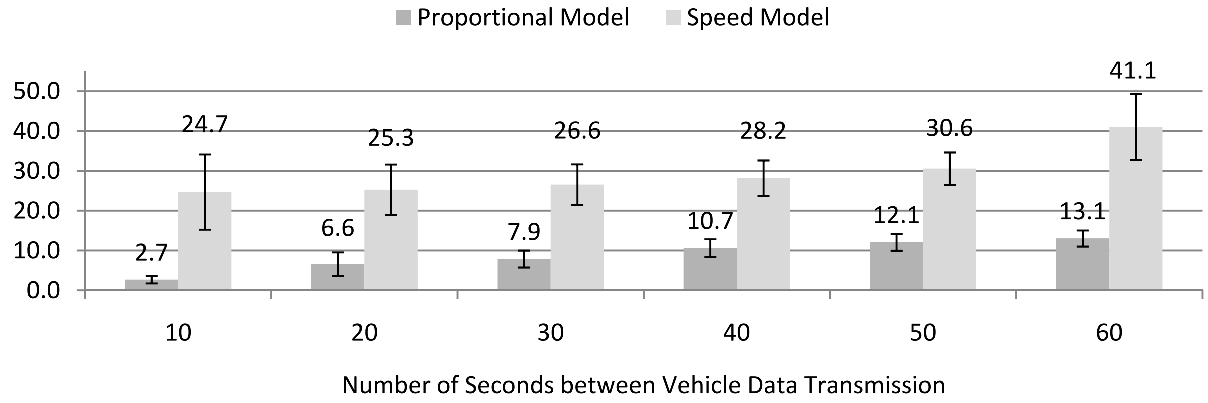
For determining the time to traverse a roadway, two different models were compared. The speed model uses all of the speed and location data that was gathered from an individual vehicle traversing a roadway that has a corresponding edge. Once the vehicle exits the edge, all of the speeds of that vehicle are averaged together. The distance of the roadway section is then divided by the average speed to produce the amount of time for the vehicle to traverse that section of the roadway. That value then becomes the weight on the edge in our transportation graph. Whenever a vehicle enters and then leaves the roadway section, the weight of the edge will be updated with that vehicle's time to traverse.

The speed model is very dependent on the speed that is transmitted from the vehicle-tracking device. The location and direction transmitted are only used to map the vehicle to an edge in the transportation graph. Since a single vehicle's data is used to determine the time to traverse the roadway, a unique identifier must be transmitted with the vehicle data to ensure the weight of the edge is updated based on the data reported by a single vehicle and not multiple vehicles that may be traveling on the same roadway section at the same time.

The proportional model for determining the time to traverse a roadway uses the data from a unique vehicle as well. Since the vehicles are transmitting their speed, location, and direction every 10 to 60 seconds, they are not necessarily transmitting this data at the start and end of an edge. If that were the case, the amount of time to traverse the edge could easily be calculated based on the times at which the data were transmitted. Since the data will be transmitted at locations other than the start and end of an edge, the proportional model will use an algorithm to proportionally determine the amount of time to traverse an edge based on the times and locations of certain data points for an individual vehicle.

The weight on the edge will be calculated after a vehicle has finished traversing the roadway section identified by an edge. The last location of the vehicle before entering the edge, up to two locations of the vehicle on the edge for which we are trying to determine the weight, and the first location of the vehicle after exiting the edge will be used to calculate the weight. As the number of locations on the edge for which we are trying to determine the weight could be zero, one, or two or more, there are three cases that must be considered. For each of the three cases, assume the following:

FIGURE II. AVERAGE DIFFERENCE IN SECONDS BETWEEN CALCULATED TIME TO TRAVERSE AND ACTUAL TIME TO TRAVERSE, INCLUDING 95% CONFIDENCE INTERVALS



1. The time to traverse from node M to node N is TTT(M, N)
2. The distance between any two locations X and Y can be found by D(X, Y)
3. The time at which a vehicle transmits the data at location X is t_X

Case 0 – 0 Locations on Edge Transmitted

For the case where no locations on the edge for which we are trying to determine the weight is transmitted, we only have the location and time for the vehicle on an edge before the edge in question and the location and time for the vehicle on an edge after the edge in question. This case is modeled in Figure I.b. Assume that location A is the last location transmitted from the vehicle before entering the edge between nodes 2 and 3, and location D is the first location transmitted from the vehicle after exiting the edge between nodes 2 and 3. The proportional model uses the following formula to calculate the weight of the edge between nodes 2 and 3:

$$\mathrm{TTT}(2, 3) = (t_D - t_A)\,\frac{D(2,3)}{D(A,D)}$$

The preceding equation finds the amount of time to traverse from the locations that were transmitted, but then multiplies it by the proportion of the distance of the edge for which we are trying to determine the weight to the total distance traveled based on the transmitted data.

Case 1 – 1 Location on Edge Transmitted

For the case where one location on the edge for which we are trying to determine the weight is transmitted, we will modify our equation from Case 0 to include the additional data point we have available. This case is modeled in Figure I.c. Assume that location A is the last location transmitted from the vehicle before entering the edge between nodes 2 and 3, location B is the one location transmitted from the vehicle while on the edge between nodes 2 and 3, and location D is the first location transmitted from the vehicle after exiting the edge between nodes 2 and 3. The proportional model uses the following formula to calculate the weight of the edge between nodes 2 and 3:

$$\mathrm{TTT}(2, 3) = \mathrm{TTT}(2, B) + \mathrm{TTT}(B, 3)$$

$$= (t_B - t_A)\,\frac{D(2,B)}{D(A,B)} + (t_D - t_B)\,\frac{D(B,3)}{D(B,D)}$$

The preceding equation breaks the edge into two different segments based on the three locations and times that were transmitted, and then determines the weight of the edge in question by multiplying the time to traverse each segment by the proportion of the edge in question that was traversed in each segment.

Case 2 – 2 or More Locations on Edge Transmitted

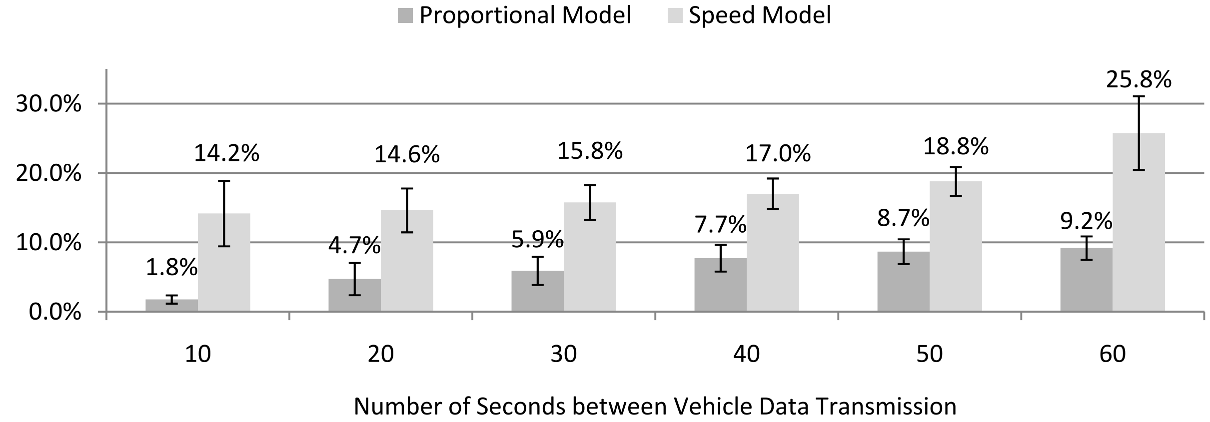
For the case where two or more locations on the edge for which we are trying to determine the weight are transmitted, we will modify our equation from Case 1 to include the additional data point we have available. This case is modeled in Figure I.d. Assume that location A is the last location transmitted from the vehicle before entering the edge between nodes 2 and 3, location B is the first location transmitted from the vehicle while on the edge between nodes 2 and 3, location C is the last location transmitted from the vehicle while on the edge between nodes 2 and 3, and location D is the first location transmitted from the vehicle after exiting the edge between nodes 2 and 3. If there are more than two locations transmitted from the vehicle while on the edge, only the first and last locations will be used in the calculation. The proportional model uses the following formula to calculate the weight of the edge between nodes 2 and 3:

$$\mathrm{TTT}(2, 3) = \mathrm{TTT}(2, B) + \mathrm{TTT}(B, C) + \mathrm{TTT}(C, 3)$$

$$= (t_B - t_A)\,\frac{D(2,B)}{D(A,B)} + (t_C - t_B) + (t_D - t_C)\,\frac{D(C,3)}{D(C,D)}$$

The preceding equation breaks the edge into three different segments based on the four locations and times that

FIGURE III. AVERAGE PERCENTAGE DIFFERENCE BETWEEN CALCULATED TIME TO TRAVERSE AND ACTUAL TIME TO TRAVERSE, INCLUDING 95% CONFIDENCE INTERVALS



were transmitted, and then determines the weight of the edge in question by multiplying the time to traverse each segment by the proportion of the edge in question that was traversed in each segment. The middle term ($t_C - t_B$) is completely on the edge in question, so the proportion of the edge in question that is traversed by that term is 1.

The results based on both the speed model and the proportional model against live data gathered from vehicles through a V2I architecture are provided in the next section.

V. SPEED MODEL AND PROPORTIONAL MODEL ANALYSIS

As described in section III, the sample roadway section used to analyze the speed model and proportional model algorithms was a roadway that experiences heavy traffic during the rush hour times of 7:00a.m.-9:00a.m. and 4:30p.m.-6:30p.m. Twenty-two probe vehicles traveled along the roadway section shown in Figure I.a and provided actual times for traveling along the roadway identified by the edge from nodes 2 to 3. While traveling along the road section, the probe vehicles were transmitting their speed, location, and direction every 10 seconds. The actual amount of time (as determined by the driver of the vehicle) was compared to the calculated time to traverse the edge between nodes 2 and 3 based on both the speed model and the proportional model.

To account for devices that do not transmit vehicle data every 10 seconds, two variables were also considered in our analysis. The amount of time between transmissions was taken as 10 seconds, 20 seconds, 30 seconds, 40 seconds, 50 seconds, and 60 seconds. Since all of the devices installed in the probe vehicles used in this study were transmitting the data every 10 seconds, to simulate 20 second data, every other data point transmitted was ignored; to simulate a vehicle transmitting data every 30 seconds, only every third data point was used in the analysis; for 40 seconds, every fourth data point was used; for 50 seconds, every fifth data point was used; and for 60 seconds, every sixth data point was used.

To further ensure an accurate simulation based on different time periods, we also considered that the first data point transmitted could change. So instead of the first data point always being the same regardless of how many seconds were assumed between transmissions, we also included that the first data point was the first, second, third, fourth, fifth, sixth, and seventh data points actually transmitted. Because the weight of the edge is dependent upon which data point is considered the last before the edge, the first on the edge, the last on the edge, and the first after the edge, changing the amount of time between transmissions and the data point to consider the first point transmitted produced varying results for each execution.

To produce an accurate analysis of the data, all of these 42 different permutations of the data were averaged across all 22 executions. The roadway section was traversed at all different times of the day, including during rush hour and during light or no traffic. The speed limit on the roadway is 45mph/72.4kph, which would take 79.2 seconds at speed limit without any slowdown due to traffic signals. The average amount of time to traverse this roadway section based on the 22 probe vehicles was 148 seconds, or 2 minutes 28 seconds.

The results for the difference in the calculated amount of time to traverse the roadway based on the speed model and proportional model compared to the actual amount of time to traverse it are provided in Figure II. Figure III provides the difference in percentage for the same comparison. Figures II and III also show the 95% confidence intervals for our data.

As can be seen, the results are as expected, with the most accurate calculations occurring when the data was transmitted in 10 second increments, and with each increasing difference between transmissions, the data gets more inaccurate.

The difference between the speed model and the proportional model is rather drastic. The speed model, even with the most granular of vehicle data transmission, has an error percentage of 14.2%, whereas the proportional model has an error percentage of only 1.8%. And although the 60 second transmission gap for the proportional model has an error percentage of 9.2%, the speed model's error percentage is rather high at 25.8%.

For the average actual time to traverse the roadway section of 148 seconds, the proportional model differed by between 2.7 and 13.1 seconds from the 10 to 60 second transmission gaps. The speed model differed by between 24.7 and 41.1 seconds from the 10 to 60 second transmission gaps.

Note that there were several statistical outliers that we decided to not include in calculating the averages for the speed model. These extremely high values skewed the averages and would have misrepresented the results. For example, one calculated data point from the speed model showed 3245 seconds for the time to travel, which has a 1854% error based on the actual time to travel of 175 seconds for that execution.

VI. CONCLUSION

In this paper, we have proposed a new algorithm for mapping discrete ITS vehicle data as gathered through a V2I architecture to determine the time to traverse roadways. Two algorithms were discussed – the speed model and the proportional model. The speed model averages all of the speeds of an individual vehicle for a roadway section and divides that value into the distance to obtain the amount of time to traverse the section. The proportional model uses location, time reported, and distance to determine the amount of time to traverse the section based on a proportional calculation of the location of vehicle data reportings to the length of the roadway section.

The study was conducted using 22 probe vehicles equipped with 10 second tracking devices where the drivers additionally timed traveling along a specific roadway. The roadway tested took an average of 148 seconds to traverse, and with 10 second data used in the calculations, the speed model differed by 24.7 seconds (14.2%), whereas the proportional model differed by only 2.7 seconds (1.8%). We conducted the same calculations with 20 second, 30 second, 40 second, 50 second, and 60 second data, and as expected, the accuracy of each model decreased with each increase in gap of the data transmission.

Although neither model produced a result with 0% error, the proportional model was extremely accurate with short transmission gaps. The error factor is due to the variability of how close to the edge start and end points the actual transmissions occurred. The closer to the intersections the transmission, the more accurate the time to traverse becomes. In addition, since we were depending on individual people to time their driving along a roadway section, there is always a small error attributable to the human factor. In addition, the error factor that is experienced with 60 second data transmissions may be acceptable based on the application that is utilizing the data.

As more vehicles become equipped with devices that are capable of transmitting vehicle data through a V2I architecture, the time to traverse multiple roadways will become more accurate. With algorithms such as the proportional model presented in this paper, even with a small percentage of vehicles equipped with tracking devices, the flow of traffic can be accurately modeled and many ITS applications developed.

VII. REFERENCES

[1] Chen, Mei. Chien I.J., Steven. "Dynamic Freeway Travel Time Prediction Using Probe Vehicle Data: Link-based vs. Path-based." *Transportation Research Board 80th Annual Meeting,* January 7-11, 2001.

[2] Miller, Jeffrey, Ellis Horowitz. "FreeSim – A V2V and V2R Freeway Traffic Simulator." *IEEE 3rd Workshop on Vehicle-to-Vehicle Communication* in conjunction with *IEEE 3rd Intelligent Vehicles Symposium*, June 2007.

[3] Pfoser, Dieter. Brakatsoulas, Sotiris. Brosch, Petra. Umlauft, Martina. Tryfona, Nektaria. Tsironis, Giorgos. "Dynamic Travel Time Provision for Road Networks." *Proceedings of the 16th ACM SIGSPATIALIinternational Conference on Advances in Geographic Information Systems,* 2008.

[4] Kothuri M., Sirisha. Tufte, Kristin. Ahn, Soyoung. Bertini L., Robert. "Using Archived ITS Data to Generate Improved Freeway Travel Time Estimates." *86th Annual Meeting of the Transportation Research Board,* January 22–26, 2007.

[5] Oh, Jun-Seok. Jayakrishnan, R. Recker, Will. "Section Travel Time Estimation from Point Detection Data." *82nd Annual Meeting of the Transportation Research Board,* January 12–16, 2003.

[6] Ding, Silin. "Freeway Travel Time Estimation using Limited Loop Data." *A Thesis Presented to The Graduate Faculty of The University of Akron,* May 2008.

[7] Cathey W., F. Dailey J., D. "A Novel Technique to Dynamically Measure Vehicle Speed using Uncalibrated Roadway Cameras." *IEEE 1st Intelligent Vehicles Symposium, 2005.*

[8] Work B., Daniel. Blandin, S´ebastien . Tossavainen, Olli-Pekka. Piccoli, Benedetto. Bayen M., Alexandre. "The PeMS algorithms for accurate, real-time estimates of g-factorsand speeds from single-loop detectors." *http://traffic.berkeley.edu /journal%20publications/WBTPB_AMRX_10.pdf*

[9] Wunnave V. Subbarao. Yen, Kang. Babij, Tadeusz. Zavaleta, Richard. Romero, Rafael. Archilla, Chrstian. "Travel Time Estimation Using Cell Phones (TTECP) for Highways and Roadways." Final Report Prepared for the Florida Department of Transportation, 2007.

[10] Work B., Daniel. Blandin, S´ebastien . Piccoli, Benedetto. Bayen M., Alexandre. "A traffic model for velocity data assimilation." *Applied Mathematics Research eXpress*, 2010.

[11] Waze. http://www.waze.com.

[12] Kong, Qing-Jie. Chen, Yikai. Liu, Yuncai. "A Fusion-Based System for Road-Network Traffic State Surveillance: A Case Study of Shanghai." IEEE Intelligent Transportation Systems Magazine, Volume 1, Number 1, Spring 2009.

[13] MIT's CarTel Project. http://cartel.csail.mit.edu.

[14] UC Berkeley's Mobile Millennium Project. http://traffic.berkeley.edu.

[15] FreeSim Web Site. http://www.freewaysimulator.com.

[16] Google Maps. http://maps.google.com.