A Sampling Theorem Approach to Traffic Sensor Optimization

W. L. Leow, Daiheng Ni, and Hossein Pishro-Nik, Members, IEEE

Abstract—With the objective of minimizing total cost that includes both sensor and congestion costs, the authors adopted a novel Sampling Theorem approach to address the problem of sensor spacing optimization. This paper presents the analysis and modeling of the power spectral density of traffic information as a two-dimensional stochastic signal using highly-detailed field data. The field data were captured by the Next Generation SIMulation program in 2005. To the best knowledge of the authors, field data with such a level of detail was previously unavailable. The resulting model enables the derivation of a characterization curve that relates sensor error to sensor spacing. The characterization curve, concurring in general with observations of a previous work, provides much more detail to facilitate sensor deployment. Based on the characterization curve and a formulation relating sensor error to congestion cost, the optimal sensor spacing that minimizes total cost can be determined.

Index Terms—Traffic sensing, sensor optimization, traffic congestion, sampling theorem, spectral domain analysis.

I. INTRODUCTION

The United States has 47,000 miles (75,640 kilometers) of Interstate Highways [1]. If half of the roads were monitored by traffic sensors with one at every one third of a mile as is typically adopted in practice, approximately 70,000 sensors will be required. Assuming each sensor has a lifetime cost of $30,000 [2], the total cost will amount to about $2 billion. If somehow, one determines that 20 percent of the sensors are unnecessary in the sense that their existence does not provide additional information, a savings of roughly $400 million is expected.

Little research has been conducted on the subject of optimal traffic sensor deployment or on the potential savings from such optimization. Current practice is mainly based on rudimentary studies or none at all. For example, Georgia Navigator [3], Georgia’s Intelligent Transportation System (ITS), chooses to install sensors at every one third of a mile along its major highways. The rationale is that this is the distance that a vehicle traverses assuming an average traffic speed of 60 miles per hour (96.6 km/hr) over a data aggregation interval of 20 seconds.

Therefore, the strategic problem that this research attempts to address is traffic sensing optimization, including optimizing sensor deployment strategies and minimizing the uncertainty of the system of interest. To achieve this goal, we sought an interdisciplinary collaboration and developed an analytical approach based on Sampling Theorem. In this approach, traffic information such as flow, speed and density is obtained from actual, detailed vehicle trajectories collected by Cambridge Systematics, Inc. under the auspices of the Next-Generation Simulation (NGSIM) program [4].

The NGSIM program has collected data sets of vehicle trajectories from actual, live traffic video footages. Advances in technology have enabled the NGSIM program to capture vehicle trajectories to a level of detail that was previously impossible. Six sets of data, each containing detailed vehicle trajectories of actual automobiles traveling on two different freeways under real-life, actual driving conditions during the morning and evening peak periods, were used in this study. The morning and evening peak periods correspond to the times of day where traffic information is of greatest importance. This study focuses on the most-important times to traffic management, instead of worst-case scenarios or extreme cases.

Signal processing techniques and the multi-dimensional Sampling Theorem will be used to glean an understanding of traffic information which is treated as a two-dimensional (2D) stochastic signal in the space-time domain. A model of the power spectral density (PSD) of traffic information as a 2D stochastic signal will then be derived. Using the derived model, we attempt to determine a characterization curve that relates sensor spacing to the error from the sensor. With this relationship, we determine the optimal sensor spacing.

Our contributions include the following:

1. A novel Sampling Theorem approach to address the optimal sensor deployment problem
2. An analytical model of the PSD of traffic information as a 2D stochastic signal
3. The normalized mean-squared error (NMSE) and sensor spacing characterization curve
4. Procedure and formulation to determine the optimal sensor spacing that minimizes the total cost.

The next section provides a review of related work. Section
III details the spectral characteristics and the modeling of traffic information and presents the NMSE and sensor spacing characterization curve. Section IV discusses the optimization of sensor spacing. This is followed by a conclusion.

II. REVIEW OF RELATED WORK

In 1979, the Federal Highway Administration (FHWA) published a guideline for locating freeway sensors [5] based on empirical studies. The report concluded that (a) any sensor spacing below 1000 ft (304.8 m) generally produces relatively little or no increase in effectiveness, (b) sensor spacing over 2500 ft (762 m) produces unsatisfactory performance according to criteria defined in the report, and (c) there exists a cost-effectiveness tradeoff for sensor spacing between 1000 ft (304.8 m) and 2500 ft (762 m). This report is similar to our research in that both address the sensor optimization problem and they lead to consistent findings. The differences between the two are (a) the report is based on empirical studies, while our approach is analytical, (b) the report employed a different set of criteria than ours to evaluate sensor deployment strategies, and (c) the report gives categorical recommendations, while our approach yields more details over the whole spectrum of sensor spacing which might be of greater interest to practitioners.

A few studies [6] [7] [8] [9] addressed the sensor location problem in a traffic system, i.e. the minimum number sensors and their associated locations to facilitate the estimation of O-D matrices. These studies differ from our research in different ways. Firstly, the objective of these studies was to estimate O-D matrices, while that of ours is to optimize sensor spacing and minimize the uncertainty of the subject system. Secondly, these studies employed strong assumptions about sensor locations (e.g. a link contains at most one sensor or sensors appear at nodes/intersections only), while in our approach a sensor may appear at any location as appropriate and a link with more than one sensor is possible. Although nodes/intersections are typical sources of traffic congestion, removing the restriction regarding sensor location is important because capacity constraints may apply at mid-block locations due to curves, grades, and accidents. Eisenman et al. [10] provided a conceptual framework of the sensor location problem and analyzed the sensitivity of the estimation and prediction quality to the number and locations of sensors. Major differences between this study and ours are the following. Firstly, rather then solving for optimal sensor locations as we do, this study assumes given sensor deployment scenarios and evaluates their effects. Secondly, from the perspective of O-D estimation, this study emphasizes high volume links where sensors are desirable, while our research deploys sensors to minimize a more general cost function including sensor costs and costs due to the loss of information. It might be interesting to find out how our work compares to these efforts. Unfortunately, a meaningful comparison has been very difficult because their stark differences in objective and approach provide no common basis for such a comparison.

Fujito et al. [11] investigated the impact of sensor spacing along freeways on the computation of performance measures. Sisiopiku [12], Thomas [13], and Oh [14] tried to find optimal loop detector locations to improve traveler information such as travel times. Woods [15] identified the need for optimizing the spacing of detectors and monitoring stations in high occupancy vehicle (HOV) lane operations. MaC Hutchon and Ryan [16] called for optimizing sensor locations for fog detection. Other applications of the sensor location problem include dilemma zone [17] and actuated signal control [18].

Applications of the sensor location problem are also found in many engineering areas other than transportation. Goulias [19] studied optimal placement of pavement temperature sensors. Stubbs and Park [20] applied Shannon's Sampling Theorem to reconstruct exact mode shapes of a structural system from a limited number of sampling points. Shi et al. [21] developed a method to prioritize sensor locations according to their ability to localize structural damage. Papadimitriou et al. [22] presented a statistical methodology for optimally locating sensors in a structure for structural model updating. Instead of monitoring structures, Berry et al. [23] studied sensor placement in municipal water networks and the model for optimizing the placement of sensors is based on mixed-integer program. Also based on mixed-integer linear program, Propato et al. [24] proposed a model identifying optimal sensor locations for water quality monitoring. Ucinski [25] studied the optimal sensor placement in a distributed system so as to maximize the accuracy of parameter identification in a 2D spatial domain.

Discussions on the Shannon Sampling Theorem and related signal processing concepts can be found in [26] [27] [28] [29]. Models for two-dimensional stochastic processes are described in [30] while [31] introduced a family of spectral density-covariance function pairs for 2D stochastic processes. We obtained several spectral density functions of demonstration models that serve as good candidates for the modeling process from [31]. Traffic flow fundamentals are covered in [32].

In summary, to our knowledge, there is no existing study that employed the same approach and accomplished the same goal as ours. More specifically, our research distinguishes itself from existing literature by the following: (a) it is an analytical approach based on Sampling Theorem, (b) it not only yields the optimal sensor spacing but also characterizes the relative merits of the entire spectrum of sensor spacing, (c) instead of restricting sensors to be at nodes or one per link as done in existing studies, our research allows sensor to be anywhere and the consideration of deploying a sensor is how it helps reduces the cost function.

III. SPECTRAL CHARACTERISTICS AND MODELING OF TRAFFIC INFORMATION

A. Theory & Spectral Characteristics of Traffic Information

Theoretical derivations based on [32-35] indicate that there exists a strong correlation between density values of
neighboring points in the space-time domain. This gives a theoretical basis for the Sampling Theorem approach. Analysis of field data using definitions and techniques in [36] further verifies the validity of this approach.

Central to this discussion is the Shannon Sampling Theorem [26], which states that for a function \( f(t) \) that contains no frequencies higher than \( W \) cycles per second, it is completely determined by giving its ordinates at a series of points spaced \( 1/2W \) s apart. The theorem can be extended to multi-dimensional signals.

In the context of the 2D traffic information signal, sampling is carried out in both the time and space domains. In the time domain, sampling rate translates to the intervals over which traffic information is aggregated. In the space domain, the sampling rate relates to the spacing between two adjacent sensors i.e. how closely the sensors are placed apart decides how close each sample of the signal is taken to the next.

In our analysis using the NGSIM data sets, we arbitrarily set the sampling rate for both time and space at \( t_{s,0} = 1/20 \). This corresponds to a data aggregation interval of 20 seconds (s) and a sensor spacing of 20 ft, which is equal to 6.096 m. It is common practice to use a data aggregation interval of 20 seconds. Too short a data aggregation interval may give rise to noisy, peaky readings while long data aggregation intervals may lead to inaccurate readings. A sensor spacing of 20 ft (6.096 m), which corresponds to the average length of a car, is inconceivable in practice but is chosen as such to sample the signal at a high frequency so as to capture the high frequency components.

In order to obtain the power spectral density (PSD) of traffic information, we first determine its 2D autocorrelation. The signal can be considered stationary since the data was collected over relatively short periods of time during the peak period and the freeway condition did not change i.e. there were no facilities like on/off ramps. The Fourier transform of the 2D autocorrelation function gives the PSD of the signal.

Magnitude plots of PSD reveals that the 2D traffic information signal has a high concentration of power in the low frequency contents. This implies that most of the spectral content can be captured using a relatively lower sampling rate, which corresponds to a larger sensor spacing.

From the PSD, it is possible to derive the NMSE associated with a particular sampling rate. The NMSE, which takes a value between zero and one, is given by

\[
\text{NMSE} = \frac{\sum_{i=1}^{N_t} \sum_{j=1}^{N_x} (X_{i,j} - \hat{X}_{i,j})^2}{\sum_{i=1}^{N_t} \sum_{j=1}^{N_x} (X_{i,j})^2} \tag{1}
\]

where

- \( N_t, N_x \) maximum index for time and space indices respectively in frequency domain
- \( X_{i,j} \) PSD magnitude at time index \( i \) and space index \( j \)
- \( \hat{X}_{i,j} \) magnitude of the truncated PSD due to lower sampling rate at \( i \) and \( j \).

In this case, the NMSE is a measure of the power loss due to spectral content not captured by the sampling i.e. spectral content that is greater than half the sampling rate.

The information carried by a signal is commensurate with the power and any loss of power can be related to information loss. By converting the sampling rate into sensor spacing, we have the means to relate NMSE to sensor spacing. A suitable model for the PSD remains to be developed in order to have a continuous characterization curve that relates NMSE to sensor spacing.

B. Modeling the PSD

We desire a model with a closed-form defining equation in order to rigorously and completely characterize the relationship between NMSE and Sensor spacing. Several demonstration models are found in [31], of which the simplest model has a spectral density function defined by

\[
S(n_1, n_2) = \frac{1}{n_1^2 + n_2^2 + 1} \tag{2}
\]

where

- \( n_1, n_2 \) are frequency-domain variables.

Plotting the spectral density function given by (2) gives a plot that has a shape similar to that given by the magnitude plot of the PSD obtained from the field data.

For the purpose of modeling, we introduce a parameter \( c \) and rewrite (2) as

\[
S(\chi, \tau, c) = \frac{c}{\chi^2 + \tau^2 + c} \tag{3}
\]

where

- \( \chi, \tau \) are frequency-domain variables that correspond to the space and time variables in the space-time domain respectively
- \( c \) is a real constant.

The presence of \( c \) in the numerator serves to normalize the model so as to facilitate the modeling process. By varying \( c \), we attempt to find a model that best fits the PSD points obtained from the NGSIM data sets. The value of \( c \) that gives the minimum root mean squared error (RMSE) was determined for each of the six NGSIM freeway data sets. We found that \( c \) ranged between 1.944 and 2.01 with a mean value of \( \bar{c} = 1.981 \).

The model defined by \( S(\chi, \tau, \bar{c}) \) is then used to characterize the variation of NMSE with sensor spacing. The complete characterization is depicted in Fig. 1. Data points from the data sets as well as the upper and lower limits of the characterization curve are also presented. The upper and lower limits corresponds to the models are given by \( S(\chi, \tau, c) \) where \( c = 1.944 \) sets the upper limit and \( c = 2.01 \) sets the lower limit.

Fig. 1 shows a good fit between the model and NGSIM data. In particular, the narrow band bounded by the upper and lower limits follows the curve closely. While it is possible that other models may fit the data equally well or even better, we opt to stay with the current model and leave the search for other, possibly better, models as a future research topic.
IV. SENSOR OPTIMIZATION

A. Discussion

With reference to Fig. 1, we observe that for any sensor spacing equal to or less than 1000 ft (304.8 m), the NMSE is less than 0.05. The stretch between 1000 ft (304.8 m) and 2500 ft (762 m) is almost linear and corresponds to the reported cost-effectiveness tradeoff region. With a sensor spacing greater than 2500 ft (762 m), the characterization curve reveals an NMSE of approximately one-third. Thus, Sampling Theorem helps us examine and explain the performance of sensors as a function of sensor spacing.

Although [5] adopted an empirical approach and used a different set of criteria, it allows a cross-comparison between findings of that study and ours. Noticeably, our research provides much more details over the whole spacing spectrum and is more practically appealing to practitioners.

B. Optimization

The report [5] also included a discussion on cost-effectiveness analysis to help practitioners decide on the most cost-effective sensor spacing. To this end, we offer a scientific, analytical approach based on Sampling Theorem to determine the optimal sensor spacing.

The objective of our approach is to minimize the total cost \( t(d) \) which includes cost to install, operate and maintain sensors as well as the cost incurred by motorists who are caught in a congestion. Hence the objective function is given by

\[
\text{Minimize } t(d) = s(d) + c(d)
\]

where

- \( s(d) \) is the annual cost of sensors that includes both capital and operating costs.
- \( c(d) \) is the annual cost of congestion due to sensor deficiency, as explained below.
- \( d \) is the sensor spacing.
- \( s_c \) is the annual cost per sensor.
- \( n \) is the number of sensors involved. \( n = (L/d+1) \) where \( L \) is the length of the road section monitored by sensors. Note that the number “1” accounts for the extra sensor needed to close the road section.

Therefore, \( s(d) \) can be expressed as:

\[
s(d) = s_c \left( \frac{L}{d} + 1 \right)
\]

The authors are not aware of any formulation that relates the NMSE to the cost incurred by motorist due to congestion. As such, a formulation based on Traffic Flow Theory was developed to define the function \( c(d) \). The formulation assumes that the freeway is managed by a traffic management center (TMC) that will not allow the Level of Service (LOS) [38] provided by the freeway lane to degrade beyond LOS D. To achieve this objective, the TMC monitors the real-time condition of the freeway using the sensors placed \( d \) distance apart and diverts traffic to alternative routes once the freeway operates at LOS D. In this formulation, congestion arises when error from the sensors causes the TMC to activate the diversion measures too late. It is also assumed that congestion occurs only during peak periods and traffic operates at LOS A at other times. During a peak period, the traffic may build up towards the conditions that define LOS D. In order to formulate the congestion cost and for the sake of simplicity, it is assumed that traffic builds up towards the peak linearly, as is depicted in Fig. 2, which presents the variation of flow \( q \) with time of the day where the service volume at LOS X is denoted by \( q_X \). The linear model, while simplifying the formulation, captures the dominant characteristic of traffic flow during peak periods i.e. a build-up towards a peak value. Other models that reasonably represent peak traffic behavior will be valid as well. The value of service volume or traffic flow at each LOS is obtained from [38].

The formulation of \( c(d) \) accounts for the cost due to congestion arising from the error from the system of sensors. The error can cause the system of sensors to either over-report or under-report the actual flow of the traffic. Over-report refers to the situation where the system of sensors gives a reading higher than the actual value and this does not cause any congestion. On the other hand, under-reporting occurs when the system gives a lower reading than the actual value i.e. \( q_{\text{reading}} = (1 - e(d)) q_{\text{actual}} \). Note that the Error, \( e(d) \), takes a value between 0 and 1. This results in traffic congestion. Due to under-reporting, congestion occurs at \( (1 - e(d)) q_D \). It is assumed...
that the sensor system is equally likely to over- or under-report. Ignoring other effects for simplicity, the shaded triangles in Fig. 2 represent the vehicles that are caught in the congestion due to the under-reporting of the sensors. The area of the shaded triangles gives the number of vehicles caught in the congestion.

With this formulation, annual cost of congestion due to sensor error is given by

\[
c(d) = N_FN_Ppc \cdot LD_p \cdot \left( \frac{d^2}{2(q_D - q_A)} \right) (e(d))^2
\]

(6)

where

- \( N_F \) number of days in a year
- \( N_P \) number of peak periods in a day
- \( D_P \) duration of peak periods (in hours)
- \( p \) probability of under-reporting
- \( c_V \) cost of congestion per vehicle mile
- \( q_S \) service volume at LOS \( X \), where \( X = A \) or \( D \)

For any deployment of sensors, denoting the random variables of the actual flow value and the sensor reading using \( Q \) and \( Q' \) respectively, we can write

\[
E[|Q-Q'|] \leq \sqrt{\text{NMSE}} (E[Q^2])^{0.5}
\]

Based on the congestion cost formulation, in the case of under-reporting when \( Q=q_D \) (a degenerate random variable),

\[
E[q_D - Q'] \leq \sqrt{\text{NMSE}} (E[q_D^2])^{0.5} \Rightarrow E[Q'] \geq (1-\sqrt{\text{NMSE}}) q_D.
\]

This gives a lower bound on the value of sensor reading when the actual flow value is \( q_D \). In other words, \( e(d) \leq \sqrt{\text{NMSE}} \). For brevity, we write NMSE instead of \( \text{NMSE}(d) \) while bearing in mind NMSE is a function of \( d \) as well. The practitioner can choose to use \( e(d) = \sqrt{\text{NMSE}} \) for the most conservative result. Alternatively, investigating a suitable function relating \( e(d) \) and \( \sqrt{\text{NMSE}} \) is a good future research topic.

For illustrative purposes, we assume \( e(d) = \text{NMSE} \leq \sqrt{\text{NMSE}} \). The inequality \( \text{NMSE} \leq \sqrt{\text{NMSE}} \) is true since \( 0 \leq \text{NMSE} \leq 1 \). Hence, (4) can be re-written as

\[
\text{Minimize } \tau(d) = s_c \left( \frac{L}{d} + 1 \right) + N_FN_Ppc \cdot LD_p \cdot \left( \frac{d^2}{2(q_D - q_A)} \right)(\text{NMSE})^{0.5}
\]

(7)

An illustrative example where actual values are substituted into (7) can be the following. The annual cost of a traffic sensor, \( s_c \), is estimated as $1600 based on [37]. The cost of congestion per vehicle, \( c_V \), is estimated as $0.26 based on [39]. The probability of under-reporting is assumed to be 0.5. The number of days can be taken as 365 and there are 2 peaks in a day. Each peak lasts for 3 hours. The service volumes at LOS A, \( q_A \), and LOS D, \( q_D \), are 700 and 1850 respectively according to [38]. A road length of one mile is assumed.

Fig. 3 presents the variation in total cost \( \tau(d) \) as a function of sensor spacing \( d \) based on (7) using the above values. The optimal sensor spacing is found to be 1086 ft (331 m). The dotted curves in Fig. 3 are the upper and lower limits of the congestion cost, which are derived in the same way as those limits in Fig. 1.

In [5], a cost-effectiveness tradeoff for sensor spacing is reported. In the proposed approach, the optimal sensor spacing for a given cost function \( s(d) \) is provided instead. This difference in output hinders any numerical comparison between the two methods. In addition, close to three decades of technological advancement and economic changes that separates the current work and [5] further makes any comparison between the two less purposeful.

V. CONCLUSION

Based on the Sampling Theorem, we analyzed the spectral characteristics of traffic information and modeled its PSD. Using the obtained model, a characterization curve that relates NMSE to sensor spacing was obtained. Using a formulation that relates sensor spacing \( d \) to cost incurred by motorists due to congestion, an optimal sensor spacing that minimizes a total cost was determined. Though the analysis presented in this paper is based on traffic data collected by existing sensors, the findings provide valuable insights to new deployment of traffic sensors.

The superiority of the proposed approach over an empirical method, such as that employed in [5], partly lies in its extensibility. Through the use of tunable parameters, different models for different circumstances e.g. lower speed limit, road curvature, etc. can be developed without the need for fresh new data for each circumstance. In addition, the current approach can be used with any type of sensor technology as long as a cost function can be determined. The possibility or ease with which an empirical method can be extended to account for changing circumstances or new sensor technology is not obvious.


W. L. Leow (M’03) received the B.Eng (Electrical) and the M.S.E.C.E degrees from the National University of Singapore and the University of Massachusetts Amherst respectively.

Daiheng Ni (M’06) is an Assistant Professor of Civil Engineering at University of Massachusetts, Amherst, U.S.A. Prior to that, he was a post-doctoral fellow at and received his Ph.D. from Georgia Institute of Technology, Atlanta, GA, U.S.A. His research interests include traffic flow theory and simulation, intelligent transportation systems, traffic sensing and information technology, and transportation logistics and optimization. His e-mail address is <ni@ecs.umass.edu> and his personal web page is <http://www.ecs.umass.edu/cee/ni/>.</p>

Hossein Pishro-Nik received the M.S. and Ph.D degrees in Electrical and Computer Engineering from the Georgia Institute of Technology in 2003 and 2005 respectively. Currently, he is an assistant professor in the Department of Electrical and Computer Engineering, University of Massachusetts Amherst. His research interests lie in the general area of communications, in particular, error control coding, communication networks, and vehicular communications. He received ECE Graduate Research Assistant Excellence Award and Center for Signal and Image Processing Outstanding Research Award from the Georgia Institute of Technology in 2005.