Optimal Sensor Placement for Both Traffic Control and Traveler Information Applications

Xuegang (Jeff) Ban*
Department of Civil and Environmental Engineering
Rensselaer Polytechnic Institute (RPI)
110 Eighth Street, room JEC 4034
Troy, NY 12180-3590
Phone: (518) 276-8043 Fax: (518) 276-4833
Email: banx@rpi.edu

Lianyu Chu
CCIT, Institute of Transportation Studies (ITS)
University of California Berkeley
Tel: (949) 824-1876 Fax: (949) 824-8385
E-mail: lchu@berkeley.edu

Ryan Herring
Department of Industrial Engineering and Operations Research
University of California, Berkeley
Phone: 510-642-5667 Fax: 510-642-0910
Email: ryanherring@berkeley.edu

JD Margulici
CCIT, Institute of Transportation Studies (ITS)
University of California, Berkeley
2105 Bancroft Way, Suite 300
Berkeley, CA 94720-383
Phone: 510-642-5929 Fax: 510-642-0910
Email: jd@calccit.org

For Presentation and Publication
88th Annual Meeting of Transportation Research Board
August 01, 2008

# of Words: 5,846
+ (1 table and 7 figures) = 2,000
TOTAL: 7,846

* *Corresponding Author*

Abstract

Traffic sensors have been deployed for decades to freeways to meet the requirements of various traffic/transportation applications, most noticeably traffic control and traveler information applications. A unique feature of traffic sensor deployment is that it is a continuous and evolving process. That is, with new applications emerge, additional sensors are usually required to be deployed to meet new requirements. This process is also sequential in nature as the new deployment has to consider existing sensors. In this article, we propose a modeling framework to capture this sequential decision-making process for traffic sensor deployment. The framework is based on the Dynamic Programming (DP) model the authors recently developed for determining optimal sensor deployment for freeway travel time estimation. We illustrate the framework using two applications: ramp metering control and travel time estimation. It is found that the proposed scheme can appropriately capture the decision-making process of traffic sensor deployment, and can generate optimal sensor placement at any stage by considering sensors that have already been deployed in the field. The model is tested using GPS-enabled cell phone data on a real-world freeway route in the San Francisco Bay Area.

1 Introduction

Intelligent Transportation Systems (ITS) applications rely on various types of data (such as traffic flow, speed, or occupancy), which are usually collected through traffic sensors. For example, freeway travel time estimation often requires speeds measured at certain locations. Traditionally, many traffic sensors were deployed in a case by case base without a systematic study on where and how many sensors to deploy to fulfill the needs of the applications [1]. Since traffic sensors are limited (usually expensive) resources, determination of best deployment strategies can help maximize the value of this resource with minimum possible cost.

Recently, optimal sensor placement for providing traveler information especially travel times has received much attention. Most of the studies in this line focused on empirical investigations, i.e. by varying location and/or spacing of existing sensors, to study how sensor spacing impacts the travel time estimation quality [3, 4, 5, 6, 7, 8, 9]. Although empirical studies can provide some insights regarding information quality and the number of sensors, they could not provide reasoning why sensors should be deployed at certain locations. More rigorous modeling on optimal sensor placement can hopefully solve this issue, but is currently sparse in the literature. Sherali et al. [10] propose a mixed-integer optimization model to determine optimal placement of vehicle identification readers for travel time estimation, although the model can only be solved approximately. Bartin et al. [11] show that the optimal sensor placement for travel time estimation can be determined by minimizing the weighted summation of speed variations of all roadway segments, each of which is associated with a sensor. A nearest neighbor (NN) algorithm was further developed

[1]One exception is the optimal sensor location problem for origin-destination matrix estimation, which has been widely studied in the literature. See for example [1, 2] and references therein

in [11]. However, the NN algorithm is not guaranteed to provide a globally optimal solution in polynomial time.

A dynamic programming (DP) model and a shortest-path based solution algorithm are proposed in [12] to solve optimal sensor placement for freeway travel time estimation. The model is based on the observation that under certain conditions sensor deployment can be conducted in a staged process, in which the decision on one stage only depends on the starting state of that stage and not on any previous stages. The DP model requires availability of vehicle trajectories, which are not widely available in current practice. Therefore it is essentially an analysis framework. However, with the advent of GPS-enabled smart-phone-based traffic monitoring, the DP model can be used with probe vehicle data, which we will illustrate by using the Mobile Century data set, presented in the last part of this article. It is further shown that the DP model can solve large scale sensor deployment problems for travel time estimation to optimality in polynomial time.

Most previous studies on optimal sensor placement focuses on single applications only (e.g. travel time estimation). [2] In reality, sensors are seldom used for single purposes. Ideally one should consider all possible applications simultaneously and generate "optimal" sensor deployment to meet requirements of all these applications. However, this is highly impractical because 1) new applications always emerge and we cannot completely predict what will happen even for the near future, and 2) many sensors have already been deployed in the field for various applications, which we have to consider when deploying additional sensors for new applications. As a result, sensor deployment in reality works in a sequential manner with sensors deployed at different stages for different applications. One example of this is that freeway loop detectors were originally deployed for traffic control purposes, mainly ramp metering control. However, as the need to generate and disseminate traveler information emerged, they are now used (and may be supplemented by new sensors as well) to produce freeway travel time estimates.

In this article, we aim to model this sequential decision process. In particular, how can we make informed (or optimal) decisions on sensor deployment for certain application given some sensors have already been deployed? We illustrate the ideas using specific traffic control and traveler information applications. For this purpose, the answers to the following questions are crucial:

(1) Suppose we have a freeway route with a number of existing sensors. Are the existing sensors sufficient for traffic control purposes? If not, how to optimally supplement existing sensors to achieve the desired control goal?

(2) If the answers to (1) are affirmative, are the sensors sufficient for providing traveler information such as estimating travel times? If not, how to optimally supplement them to achieve desired quality of traveler information?

Notice that we group the questions as first for traffic control and then for traveler information applications. This is because we believe that these two types of applications have different priorities. At least from traffic management point of view, effective and efficient traffic management and control is the first priority. This is due to the following two reasons. First, historically traffic control and management is the focus of most traffic management agencies (like DOTs). This is true even when traveler information is becoming more crucial nowadays. Second and more importantly, a well managed and operated transportation system is more predictable and is thus the basis of effective traveler information. It is hard to imagine that traveler information will have significant value on a poorly managed transportation system. Therefore, we need to solve sensor placement for traffic control applications first. This is actually what is happening now: there

[2] One exception is Eisenman et al. [13] who provide an information learning based conceptual framework of sensor placement for various applications. But the framework is too general to be practically implementable.

are already some sensors in place (most likely for some traffic control purposes) and we need to optimally enhance these sensors for new applications (e.g. traveler information applications).

In this article, we try to answer the above two questions in a sequential yet coherent manner. In particular, we focus on two applications: freeway ramp metering control and providing freeway travel times. The key is the modeling ability to optimally add additional sensors to the field if needed to meet the requirements of new applications. In this article, we show that the DP model developed in [12] can be extended to determine optimal sensor placement for other applications such as occupancy estimation. Furthermore, it is able to consider existing sensors when determining the optimal locations of new sensors. In Section 2, the DP model and solution algorithm is briefly described, together with the method of how to optimally incorporate existing sensors. The ramp metering control application is discussed in Section 3. In this article, we focus on SWARM [14] which requires freeway mainline occupancy as the major input. SWARM also requires sensors at fixed locations (usually upstream) for each on-ramp, and we will then show how to determine the optimal locations of additional sensors to have appropriate estimation of freeway occupancy. Travel time application is presented in Section 4. Section 5 focuses on numerical examples based on a segment of I-880 in the San Francisco Bay Area. The data are obtained from a field experiment which deployed for 10 hours 100 cars equipped with GPS-enabled cellular phones to collect traffic data along the segment of freeway. We conclude our work in Section 6.

2 Dynamic Programming Model for Optimal Sensor Placement

2.1 A DP Model

Denote the study freeway segment as route r with length L. Assume the duration of the study period is T. WSe discretize time into *intervals* and space into *sections*. Each interval has fixed duration Δt such as 30 seconds; each section has also fixed length Δx such as 50 or 100 feet. We assume $L = N\Delta x$ and $T = H\Delta t$, i.e. we have in total N sections and H time intervals. Given this setting, suppose we are interested in some generic traffic measurement which could be for example speed, occupancy, etc. In particular, we denote the *ground-truth* measurement at time t ($1 \le t \le H$) and section n ($1 \le n \le N$) as $u(n,t)$. Assume we are to deploy K sensors to route r and in general $K \ll N$. Further denote the *estimated* measurement at time t and section n is $\bar{u}(n,t)$, which is generated by the sensors. Here we assume that sensors should be deployed to minimize the deviation between the ground-truth and estimated measurements. A commonly used metric is the mean square error (MSE), defined as follows:

$$E = \frac{\sum_{n=1}^{N}\sum_{t=1}^{H}(u(n,t) - \bar{u}(n,t))^2}{NH}. \tag{1}$$

In this article, E is the objective function that will be used to determine the optimal sensor placement. Now the question is how the estimated measurements are obtained given sensor placement. We adopt a simple yet practical scheme: each sensor is associated with a spatial *influence area*, called a link. Each link contains one or multiple sections and the link boundaries are the section boundaries. We assume the estimated measurement of every section within a link is identical to the measurement collected by the sensor associated with the link. This is illustrated in Figure 1, which depicts a route with 7 sections numbered 1 - 7. The three sensors are denoted as I, II, III. Sensor I is associated with a link that contains sections 1-3, the link of sensor II contains section 4, and sections 5-7 are for the link associated with sensor III. The ground-truth and estimated measurements for a given time t are shown using two curves at the top of Figure 1. Notice that since sensor I is deployed at section 2, the sensor-generated measurement by I is $u(2,t)$ for any time interval t. In other words, we assume in this article that sensors are "perfect" in detecting the measurement. As a result, the estimated measurements for the first three sections are $\bar{u}(i,t) = u(2,t), \forall 1 \leq i \leq 3$. Similarly, we have $\bar{u}(4,t) = u(4,t)$ and $\bar{u}(i,t) = u(6,t), \forall 5 \leq i \leq 7$. That is, the estimated measurement at one section at time t is the ground-truth measurement at another section (maybe itself) at the same time. The resulting estimated measurements are represented using the step-wise curve since we assume the sensor measurement is uniform across its associated link. The optimal sensor placement should then minimize the deviation of the two curves over all time intervals.

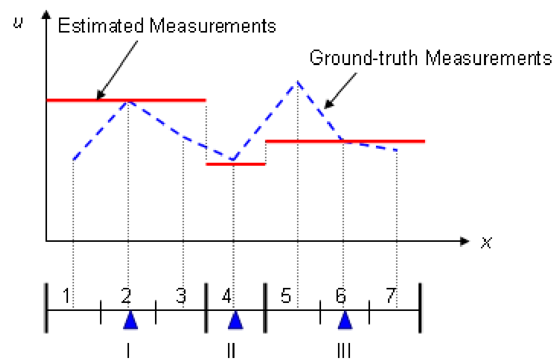


Figure 1: Calculation of Estimated Measurements

If we further assume that a sensor is always in the middle of its associated link, we can then convert the optimal sensor placement problem to a problem that aims to determine the optimal link starting and ending locations. Clearly the starting and ending locations of link k $(1 \leq k \leq K)$ coincide with sections, denoted as s_k and y_k respectively. The objective function E in (1) can then be rewritten as:

$$E = \frac{\sum_{k=1}^{K} \sum_{s_k \leq n \leq y_k} \sum_{t=1}^{H} (u(n,t) - \bar{u}(n,t))^2}{NH} = \sum_{k=1}^{K} E_k(s_k, y_k). \tag{2}$$

In the above equation, E_k is the MSE of measurements for link k which is a function of the starting and ending link locations s_k, y_k only. Specifically, E_k can be defined as:

$$E_k(s_k, y_k) = \frac{\sum_{s_k \leq n \leq y_k} \sum_{t=1}^{H} (u(n,t) - \bar{u}(n,t))^2}{NH}. \tag{3}$$

We assume a set of vehicle trajectories are available for the study route from which $u(n,t)$ can be estimated for any (n,t) pair. Under this assumption, $E_k(s_k, y_k)$ is computable for any given starting and ending location s_k, y_k. As a result, the optimal values of $s_k, y_k, 1 \leq k \leq K$ can be obtained via solving the following integer programming problem:

$$\min_{1\leq s_k,y_k\leq N,k=1,\ldots,K} \quad \sum_{k=1}^{K} E_k(s_k,y_k). \tag{4}$$

$$\text{s.t.} \qquad s_1 = 1 \tag{5}$$

$$y_K = N. \tag{6}$$

$$s_{k+1} = y_k + 1. \tag{7}$$

$$k \leq s_k \leq y_k \leq N - K + k. \tag{8}$$

Here (5) - (8) are constraints for sensor placement. Equations (5) and (6) hold because the first link must start at section 1 and the last link (link K) must end at section N. Equation (7) holds since knowing the ending section of link k (y_k), the starting section of link $(k+1)$ must be the next section (y_k+1). This is called *state transfer*. The first inequality of (8) holds since there are $k-1$ links before link k, which contain at least $k-1$ sections. Similarly, the last inequality of (8) holds since there are $K-k$ links after link k, which contain at least $K-k$ sections.

Solving the above integer programming problem for large scale problems is not tractable. In this article, we convert the problem to a DP model by 1) dividing the problem into K stages (one sensor is deployed in each stage), 2) defining s_k as the state variable and y_k as the decision variable of stage k, and 3) assuming the cost at each stage is the link MSE (E_k as defined in (3)). Under this setting, it is clear to see that the decision at stage k (i.e. the value of y_k) is only determined by the state variable of the stage (i.e. s_k) since the objective to deploy sensor k at this stage is to minimize E_k which is a function of s_k and y_k only. This observation leads to a DP formulation of the problem. Further define $F_k(s_k)$ as the total cost from stage k (including stage k) to the last stage (i.e. stage K). Then a recursive formulation exists for $F_k(s_k)$ as follows:

$$F_1(s_1) = F_1(1) = \min_{1\leq y_1\leq N-K+1} \left\{E_1(1,y_1) + F_2(y_1+1)\right\}, \tag{9}$$

$$F_k(s_k) = \min_{s_k\leq y_k\leq N-K+k} \left\{E_k(s_k,y_k) + F_{k+1}(y_k+1)\right\}, 2 \leq k \leq K-1, \tag{10}$$

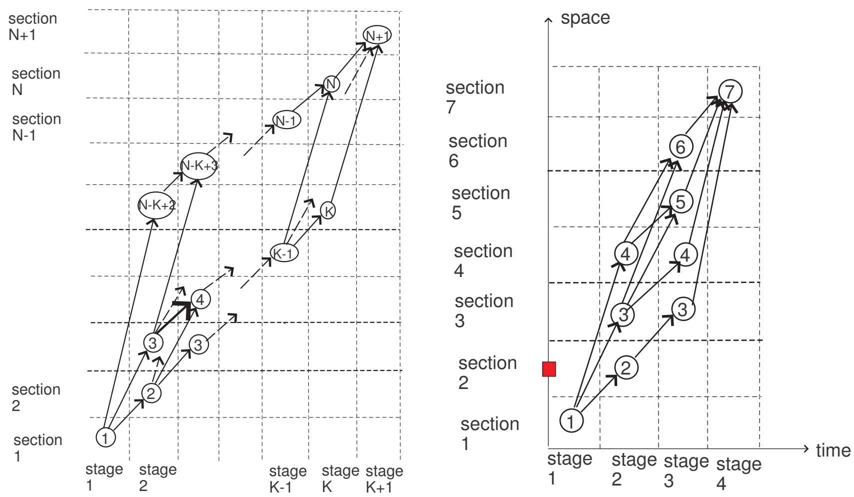
$$F_K(s_K) = E_k(s_K,N). \tag{11}$$

It can be shown that solving (4) - (8) is equivalent to solve these three recursive equations which satisfy the *optimality principle* of DP [12].

2.2 A Graph-Based Solution Algorithm

The DP model (9) - (11) can be represented as a graph shown in Figure 2(a). In the figure, stages are listed horizontally and sections are listed vertically. The state of a stage represents the starting section of the link associated with the stage. In this figure, all possible states of a stage are represented as *nodes*. In other words, a node represents a section of the roadway, and the node number is the section number. For example, the node at stage 2 and Section 2 represents that the starting location of link 2 could be section 2. As mentioned before (especially equations (5) - (8)), there is only one state in stage 1 ($s_1 = 1$) and $(N-K+1)$ states (from k to $N-K+k$) for stage $k = 2,\ldots,K$. We further create a fake stage as stage $K+1$ that has only one fake state $N+1$.

A connection, denoted as an *arc*, may be created from a node in stage k to another node in the immediate next stage $k+1$ if the latter node has a higher node number. Each arc actually represents a possible roadway link by defining the link's starting and ending sections. That is, an



(a) Graph Representation of DP Model. (b) Consideration of Existing Sensors

Figure 2: The DP Model

arc from node s_k in stage k to node s_{k+1} in stage $k+1$ represents one possible configuration for link k: it starts at section s_k and ends at section $s_{k+1} - 1$ because the next link starts at s_{k+1}. Therefore, we must have $s_{k+1} > s_k$ in order to construct the arc. For example, the arc from node 2 in stage 2 to node 4 in stage 3 (marked in bold line) in Figure 2(a) means that one possible configuration for link 2: it starts at node 2 and ends at node 3 (both are inclusive). Therefore, there should be no arc from node 4 in stage 2 to node 4 or lower in stage 3. Furthermore, there are no arcs between any two stages that are not adjacent to each other. We also associate a cost with each arc in Figure 2(a). For the arc from node s_k in stage k to node s_{k+1} in stage $k+1$, the arc cost is $E_k(s_k, s_{k+1} - 1)$ as computed in (3). In other words, the cost of an arc is the MSE of travel time estimation for its corresponding roadway link.

It is easy to check that the graph constructed in the above manner enumerates all possible states in each stage (1 to K) and all possible configurations (i.e., the starting and ending locations) of each link. It also incorporates all the constraints of the model shown in equations (5) - (8). More importantly, each path from node 1 in stage 1 to node $N+1$ in stage $K+1$ contains exactly K arcs, each of which represents a possible configuration of a particular roadway link (i.e. its starting and ending sections). In other words, each path represents a potential sensor deployment scenario. Therefore the optimal sensor locations can be achieved by finding the minimum-cost path from node 1 in stage 1 to node $N+1$ in stage $K+1$. Since all arc costs are positive, the DP model can be solved by a shortest-path search algorithm. Furthermore, the time complexity of solving the DP is $O(KN^2)$ if $N \gg K \gg 2$, which is polynomial.

2.3 Consideration of Existing Sensors

One feature of the DP model and its graph-based solution approach is that existing sensors can be easily considered. In this case, we make a simple adjustment to the dynamic programming graph representation of the solution space. First, we match all existing sensors to the appropriate section they reside in. Then, every possible link (represented as an arc in the graph) that covers a section with an existing sensor in it but does not have the existing sensor at the center of the link is removed from consideration as a possible choice in the final solution. The reason for this is that we assume a sensor must be in the middle of its associated link.

As an example, Figure 2(b) shows a highway segment that is broken down into 6 sections. Suppose that we already have a sensor in section 2. If this is the case, then we cannot consider links that cover section 2 but do not have section 2 as the middle of the link. This means that a link covering sections 1 through 4 would not be permissible in the solution (because that would imply a sensor in 3 and not on section 2 based on the fact that a sensor must be in the middle of its associated link). This also implies that the arc from node 1 in stage 1 to node 5 in stage 2 in the DP graph should be eliminated. On the other hand, a link covering sections 1 through 3 would be permissible, implying that the arc from node 1 in stage 1 to node 4 in stage 2 should be included. This is also the case for the arc from node 1 in stage 1 to node 2 in stage 2, and the arc from node 1 in stage 1 to node 3 in stage 2. Furthermore, the arcs from node 2 in stage 2 to nodes 4, 5, and 6 in stage 3 should all be eliminated. The graph in Figure 2(b) shows the adjusted DP graph after removing all impermissible links.

As a result, to account for existing sensors, one can use a simple linear search on all of the links to identify which ones to remove, and then uses the shortest path algorithm described in Section 2.2 to compute the final solution on the adjusted graph. Therefore, the complexity of the algorithm remains the same as the original DP algorithm, i.e. $O(KN^2)$ for $N \gg K \gg 2$. The DP model and graph-based solution technique presented in this section is general and may be applied to freeway speeds, occupancies, and travel times. The only difference for different applications is how the objective function is defined, especially the link MSE E_k as defined in (3). In the next two sections, we discuss how the model can be used for ramp metering and freeway travel time applications.

3 Optimal Sensor Placement for Ramp Metering

3.1 Ramp Metering Background and History

Ramp metering has been recognized as an effective freeway management strategy to avoid or ameliorate freeway traffic congestion by limiting access to the freeway. The benefits of ramp metering are:

(1) Restrict vehicles entering freeway by temporarily storing them on the ramps in order to ensure that mainline freeway is operated within capacity and thus prevent congestion.

(2) Break up platoons of vehicles entering freeways in order for vehicles from onramps to merge to the freeway mainline more easily and thus enhance safety.

(3) Divert vehicles that cannot afford waiting on the onramps to other routes and thus reduce demand to the freeway.

In practice, methods of metering operation can be divided into two primary categories: fixed-time (or pre-timed) control and adaptive (or traffic responsive) control. Pre-timed control utilizes Time-of-Day metering rates that are pre-determined to best manage "expected" conditions based on an analysis of historical data. Adaptive control dynamically modifies metering rates based on real-time traffic data, thus, conceptually allowing for better responses to variations in traffic conditions.

The adaptive or traffic responsive ramp metering control can be further classified as local traffic responsive control and coordinated traffic responsive control. Local traffic responsive control determines metering rates based on current prevailing mainline traffic conditions in the vicinity of the ramp. Examples are demand-capacity control, occupancy control, and feedback control [15]. Coordinated traffic responsive control determines metering rates based on the prevailing traffic conditions of an extended section of roadway. Notable instances include ZONE in Minnesota, BOTTLENECK in Washington, and SWARM in California and Oregon [16, 17].

3.2 Sensor and Data Requirement for Ramp Metering

Local traffic-responsive ramp metering control needs to obtain traffic condition data from sensors on the freeway mainline. Typically, these sensors are required to be placed upstream of the ramp. This requirement applies to many ramp metering systems deployed in the real world, such as the three metering systems in California: Semi-Actuated Traffic Management System (SATMS), San Diego Ramp Metering System (SDRMS), and Traffic Operations System (TOS). All three systems need mainline traffic flow and occupancy data to operate.

Coordinated traffic-responsive ramp metering control seeks to optimize a multiple-ramp section of a highway, often with the control of flow or occupancy through a bottleneck as the ultimate goal. Typically, sensors are needed to be placed to the mainline freeway evenly at a certain space and/or at bottleneck locations. For example, ZONE needs volume data; BOTTLENECK needs occupancy data; SWARM needs either volume and/or occupancy data.

By looking at the type of data requirement, it is found that occupancy data are widely used and required by most ramp metering algorithms. In this article, we focus on SWARM and investigate how sensors can be deployed to better facilitate this metering algorithm.

3.3 Optimal Sensor Placement for SWARM

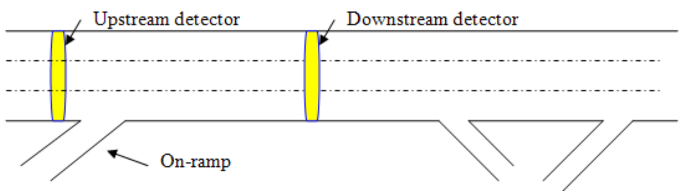
SWARM is a system-level ramp metering system. It is operated as a central ramp metering system at Traffic Management Center (TMC). SWARM has four algorithms, SWARM 1, SWARM 2a, SWARM 2b, and SWARM 2c. A meter operated under SWARM can be set up to use either of them or the combination of them. SWARM needs a mainline sensors to be placed upstream of each onramp. In order for SWARM 2b and 2c to work appropriately, as shown in Figure 3(a), SWARM requires a sensor located upstream of the ramp and another located downstream of the ramp. The mainline sensor upstream of the next onramp can be used as the "downstream" sensor, as shown in Figure 3(b). However, this relaxation will not work well if thee is a bottleneck, caused by either lane drop or strong weaving or merging, between the two mainline upstream sensor. As a result, we assume that the upstream sensors of all onramps as given (since they have to be deployed as a requirement by the algorithm); the downstream sensors however need to be "optimally" determined in terms of both numbers and actual locations to better estimate the mainline occupancy. As discussed in Section 2, the objective is to minimize the deviation of ground-truth and estimated mainline occupancy, as defined as follows:

$$E_c = \sum_{k=1}^{K} E_k^c(s_k, y_k). \tag{12}$$

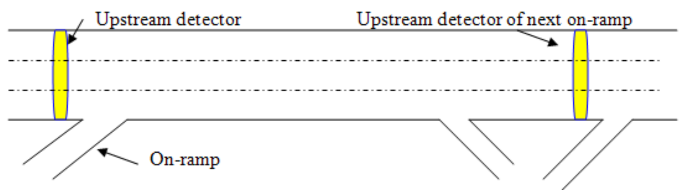
Here E_c denotes the objective function for ramp metering (i.e. for occupancy) and E_k^c is the MSE of occupancy for link k which can be defined as:

$$E_k^c(s_k, y_k) = \frac{\sum_{s_k \le n \le y_k} \sum_{t=1}^{H} (o(n,t) - \bar{o}(n,t))^2}{NH}. \tag{13}$$

We use $o(n,t)$ and $\bar{o}(n,t)$ to denote respectively the ground truth and estimated occupancies at section n and time t.



(a) Normal Sensor requirement for SWARM 2c



(b) Relaxed Sensor requirement for SWARM 2c

Figure 3: Sensor Placement Requirement for SWARM Ramp Metering Algorithm

4 Optimal Sensor Placement for Freeway Travel Time Estimation

Detailed discussions on how to apply the DP model to determine optimal sensor deployment for freeway travel time estimation are provided in [12]. In this section, we briefly discuss how the objective function is defined. For this purpose, we first denote respectively $\hat{\tau}_k^m$ and τ_k^m the estimated and actual travel times of the m-th vehicle $(1 \le m \le M)$ traveling link k $(1 \le k \le K)$, and M is the total number of vehicles. The travel time estimation error for the m-th vehicle on link k, denoted as e_k^m, can be expressed as:

$$e_k^m = \hat{\tau}_k^m - \tau_k^m. \tag{14}$$

We then use the same objective function as that in [11, 12], which is defined as follows:

$$E_t = \frac{\sum_{m=1}^{M} \sum_{k=1}^{K} (e_k^m)^2}{M} = \sum_{k=1}^{K} E_k^t. \tag{15}$$

Here E_t represents the objective function for travel time estimation. E_k^t is the Mean Square Error (MSE) of the travel time estimation for all M vehicles for link k, defined as:

$$E_k^t = \frac{\sum_{m=1}^{M} (e_k^m)^2}{M}. \tag{16}$$

The objective defined in (15) focuses on estimation errors of all individual links, instead of only on the entire route. The reason for this is that we want to generate sensor locations that can provide "good" estimates for all link travel times, not only in terms of the entire route. If attention is only put on the entire route, it is possible that the resulting sensor locations may underestimate travel times for certain links and overestimate for other links; but as a whole, they cancel out each other and provide good estimation. This type of sensor placement is not desirable. It is easy to see that the objective function we use here can effectively eliminate such sensor deployment strategies since they will lead to large objective values using equation (15). In [12], it is shown that this objective function definition is effective to generate sensor placement that is optimal to both the (entire) route r and its sub-routes.

5 Case Studies

We present a case study in this section to illustrate how the optimal sensor placement can be determined by considering ramp metering and travel time estimation in a sequential manner. The case study is based on Mobile Century data. Mobile Century is an experiment performed on February 8th, 2008, in which 165 drivers drove 100 vehicles on Interstate 880 (see Figure 4) for 10 hours in loops of length 5.5 to 10 miles [18, 19]. The experiment involved each vehicle carrying a Nokia N95 GPS-enabled smartphone, transmitting in real time loop detector-like data (called VTL data for Virtual Trip Line), which consists of speed readings at GPS-defined locations upon crossing of the locations. These VTLs represent "virtual" loop detectors, which are smartphone-based and may be used by phone manufacturers and access providers to monitor traffic in the near future. The experiment achieved a 2% to 5% penetration rate on the highway throughout the day, thus mimicking smartphone penetration in the driving population in about 18 months. In addition to this online transmitted data, each of the GPS logs collected by the phones at a 1/3 Hz rate was saved in the memory of the phone. While using trajectory data is not part of the Mobile Century technology development plan, this archival data collected from the experiment can be of great use for traffic modeling and analysis (as will be shown later in this section).

We select the shorter loop from CA-84 (postmile 20) to Tennyson St (post mile 25.5) as the study site in this article. The two circles in Figure 4 show the starting and ending locations of the route, which is about 5.5 miles. There are five major interchanges along this route at Decoto Rd, Alvarado Blvd, Alvarado Niles Rd, Wipple Rd, and Industrial Pkwy respectively. All the on-ramps to I-880 at these five locations are metered. The currently deployed ramp metering strategy is TOS, which is local responsive. TOS may be upgraded in the future to system-wide metering strategies. In this article, we select SWARM as one alternative for system-wide metering scheme. Since upstream sensors are required by SWARM (see Section 3.3), we consider the nearest upstream mainline sensor at each major interchange as existing sensors. Their locations are marked using solid lines in Figure 4 together with their exact location in postmile.

If we evaluate equation (2) for the five existing sensors for occupancy, the resulting objective value (we take the square root of (2) hereafter in this article so that the objective value has the same unit as the traffic measurement, i.e. occupancy or travel time) is 0.0264 (2.64%). This implies that the average deviation between the ground-truth occupancy (calculated using vehicle trajectories from Mobile Century) and the estimated occupancy (estimated using the scheme in Section 2.1) is 2.64%. For illustration purpose, suppose we require the deviation must be less than 0.0225 (2.25%). We will then need to add more sensors to this route. As shown in Section 2.3, the optimal placement of additional sensors can be determined using the DP model. The result shows that adding 4 additional sensors will service the purpose and the deviation is reduced to 0.0224 (2.24%). Figure 5 depicts (using the line marked as "Occ") how the occupancy objective

Figure 4: Study Site of the Case Study (Source: maps.google.com)

value changes as the number of sensors increases from 5 to 9. The reduction of the objective value is monotonic (meaning the estimation quality is improved with more sensors deployed) which is a desirable feature of the DP algorithm.

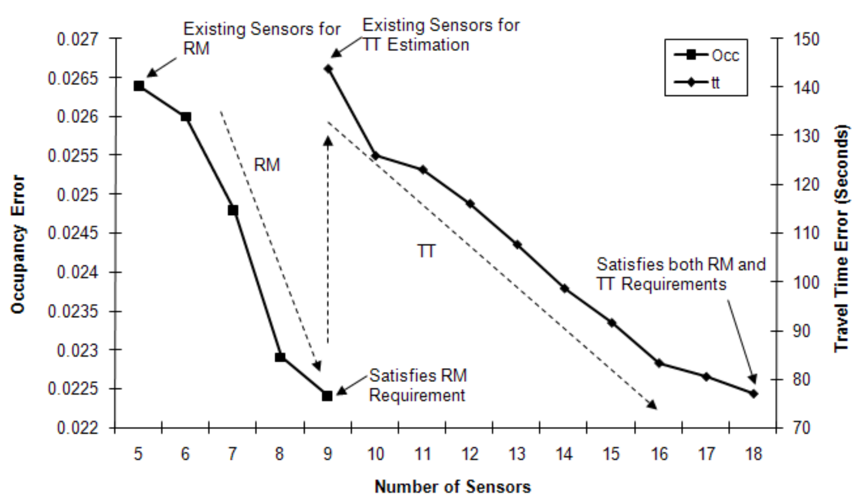


Figure 5: Change of Objective Values vs. Number of Sensors

After satisfying the ramp metering application requirement, we now focus on the travel time estimation application. The nine sensors produce an objective value of 143.8 seconds based on equation (16). If we require the objective value for travel time must be less than 80 seconds (which represents about 13% error since the travel time of the entire route is about 10 minutes, i.e. 600 seconds). It turns out that 9 additional sensors are needed, resulting in 18 sensors in total for this segment. The line marked with "tt" in Figure 5 illustrates how the travel time objective value changes as the number of sensors increases from 9 (as a result of ramp metering requirement) to 18. Again, the decrease of the travel time estimation error is monotonic. We show the actual locations of the 18 sensors in Table 1. The second column of the table lists the exact postmiles of the 18 sensors, while columns 3-5 indicate using "$\surd$" whether a particular sensor is considered as existing or generated for ramp metering or for travel time estimation.

Sensor Index	PM	Existing Sensors	Ramp Metering	Travel Time
1	20.51	$\surd$		
2	21.28	$\surd$		
3	22.1		$\surd$	
4	22.98			$\surd$
5	23.33	$\surd$		
6	23.68			$\surd$
7	24.01	$\surd$		
8	24.12			$\surd$
9	24.24		$\surd$	
10	24.31			$\surd$
11	24.4			$\surd$
12	24.48	$\surd$		
13	24.57			$\surd$
14	24.71			$\surd$
15	24.8		$\surd$	
16	25.01			$\surd$
17	25.3		$\surd$	
18	25.45			$\surd$

Table 1: Optimal Sensor Locations for the Case Study

The dashed arrows in Figure 5 depict how the objective values for the two applications change as the number of sensors increases. Notice that we first focus on ramp metering application (as marked as "RM") which has higher priority. As we reach the stage when we have 9 sensors, the ramp metering requirement is met. This is the time when we switch our focus to travel time estimation. The vertical dashed arrow indicates this switch and the 9 sensors from the ramp metering application is the "existing" sensors for travel time estimation. After the switch, we concentrate on the travel time application as marked by "TT." Notice that finally we deploy 18 sensors, which satisfy the requirements for both ramp metering and travel time estimation applications. It is the authors' understanding that the above decision-making process is more closely related to what is happening in practice.

To show whether the generated sensor placement makes sense, we associate the sensor locations with the speed contour map of the route. Figure 6 depicts the speed contour map of the route with darker color representing more congested areas. It is clear that the major congestion area is roughly from PM 23.5 to 25.5. The triangles on the y-axis of the figure shows the sensor locations for 9 sensors. By comparing the locations of existing five sensors in Figure 4 (also shown in Table 1), we can see that only 1 additional sensor is deployed to the free flow area (at PM 22.1) and the other 3 sensors are all deployed to the congestion area (at PM 24.24, 24.8, and 25.3 respectively). This intuitively makes sense since traffic conditions at congestion areas are usually more complicated which need to be captured by additional sensors.

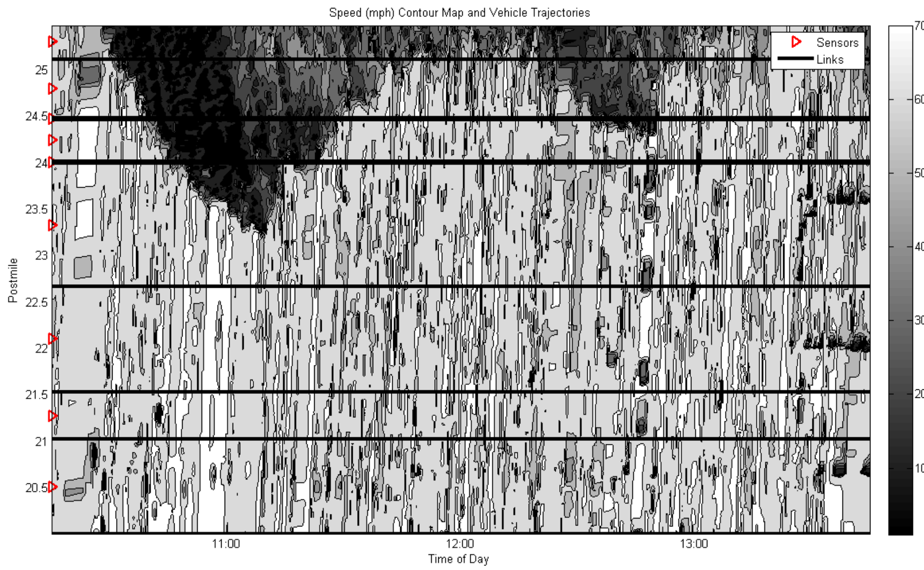


Figure 6: Optimal Sensor Placement vs. Speed Contour Map (9 Sensors)

We further depict in Figure 7 the generated sensor locations by the DP model as we go through the entire process, i.e. by increasing the number of sensors from 5 (the beginning state) to 18 (which satisfies both the ramp metering and travel time estimation). In the figure, asterisks represent the locations of sensors. We can see that as more sensors are deployed, most of them are deployed to the congestion area (PM 23.5 to 25.5); only 1 (up to 11 sensors in total) or 2 (12 sensors or above in total) additional sensors are deployed into the free flow area. More importantly, as additional sensors are added in, the locations of previously deployed sensors in congestion areas remain almost unchanged. This is illustrated using the solid thin lines in the figure (dashed lines indicate the locations of the five existing sensors), which show that locations of newly deployed sensors just "branch out" from existing sensors in congestion areas. This implies that the DP algorithm has the ability to capture the most significant congestion area and if more sensors are available, the second most significant congestion area will be captured and so on. The locations of sensors in free flow areas however may change since the speeds detected in free flow areas are not sensitive to the actual sensor locations. The above discussions illustrate the close correlation between the optimal sensor locations generated by the DP algorithm and the congestion areas of the network. They also show that the results from DP are stable and predictable, which is desirable in practice.

To illustrate that the generated optimal sensor placement is superior to that generated by purely engineering judgement, we compare the performance of the model-generated sensor placement with sensors that have already deployed in the field along the study route. As shown on PeMS (Performance Measurement Systems, [20]), there are 13 sensors deployed on I-880 NB for the study route (i.e. from PM 20 to 25.5). By evaluating the objective value of these 13 sensors using equation (15), we can obtain that the estimation error is about 159.50 seconds, roughly 26.6% (the route travel time is 10 minutes). Now we assume we have these 13 sensors but will place them at different locations based on the DP model. This results in an estimation error of 107.70 seconds, approximately 18.0%, which can be seen from Figure 5. In other words, the modeling framework proposed in this article could potentially improve the travel time estimation quality by 8.6% by having installed those sensors at more appropriate locations, and without any other extra cost. We notice that this comparison only considers ramp metering and travel time estimation. However, it

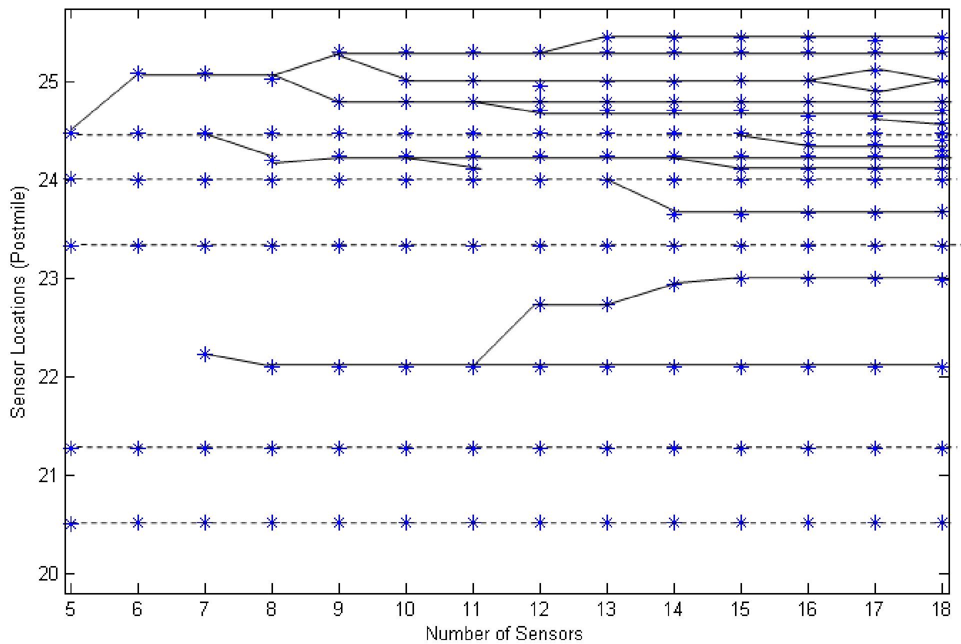


Figure 7: Evolution of Optimal Sensor Locations (from 5 to 18 Sensors)

is evident from the comparison that the benefit of using the proposed framework is significant.

6 Conclusion

We proposed a modeling framework to study optimal traffic sensor placement. The proposed method aims to capture the fact that traffic sensor deployment in reality is a continuous, evolving, and sequential process, i.e. the placement of additional sensors for new applications has to consider sensors that have already been deployed. In this article, we adopted the Dynamic Programming (DP) modeling framework recently developed by the authors for optimal sensor deployment, which has the capability to optimally deploy additional sensors by considering existing sensors. For two traffic applications: ramp metering and travel time estimation, we showed, using the Mobile Century GPS data on a real-world freeway route, that the proposed framework can generate optimal sensor placement for both applications in a sequential yet coherent manner. The generated optimal sensor placement matches well with the congestion areas of the study route, which further shows that the placement is reasonable. The optimal placement generated by the model is also superior to existing sensor placement in terms of providing better estimates to freeway travel times.

The proposed method and algorithm need to be further tested and validated on more and larger real-world traffic applications. Research in this direction will be pursued in future work and results will be reported in subsequent articles.

References

[1] H. Yang and J. Zhou. Optimal traffic ccounting locations for origin-destination matrix estimation. *Transportation Research B*, 32(1):109–126, 1998.

[2] L. Bianco, G. Confessore, and P. Reverberi. A network based model for traffic sensor location with implications on O-D matrix estimates. *Transportation Science*, 35(1):50–60, 2001.

[3] K. Ozbay, B. Bartin, and S. Chien. South Jersey real-time motorist information systems: Technology and practice. *Transportation Research Record*, 1886:68–75, 2004.

[4] I. Fujito, R. Margiotta, W. Huang, and W.A. Perez. The effect of sensor spacing on performance measure calculations. In *Proceedings of the 85th Annual Meeting of Transportation Research Board (CD-ROM)*, 2006.

[5] J. Kwon, B. McCullough, K. Petty, and P. Varaiya. Evaluation of PeMS to improve the congestion monitoring program. Technical report, Final Report for PATH TO 5319, 2006.

[6] X. Ban, Y. Li, A. Skabardonis, and J.D. Margulici. Performance evaluation of travel time methods for real time traffic applications. In *Proceedings of the 11th World Congress on Transport Research (CD-ROM)*, 2007.

[7] G. Thomas. The relationship between detector location and travel characteristics on arterial streets. *Institute of Transportation Engineers Journal*, 69(10):36–42, 1999.

[8] S. Oh, B. Ran, and K. Choi. Optimal detector location for estimating link travel time speed in urban arterial roads. In *Proceedings of the 82nd Annual Meetings of the Transportation Research Board (CD-ROM)*, 2003.

[9] S. Jung, A. Toppen, and K. Wunderlich. The effect of average loop detector spacing on the accuracy of calculated travel times: Twin cities case study. Technical report, Mitretek Systems,, 2005.

[10] H.D. Sherali, J. Desai, and H. Rakha. A discrete optimization approach for locating automatic vehicle identification readers for the provision of roadway travel times. *Transportation Research B*, 40:857–871, 2006.

[11] B. Bartin, K. Ozbay, and C. Iyigun. A clustering based methodology for determining the optimal roadway configuration of detectors for travel time estimation. *Transportation Research Record*, 2000:98–105, 2007.

[12] X. Ban, R. Herring, JD Margulici, and A. Bayen. Optimal sensor placement for freeway travel time estimation. *Submitted to the 18th International Symposium on Transportation and Traffic Theory*, 2008.

[13] S.M. Eisenman, X. Fei, X. Zhou, and H.S. Mahmassani. Number and location of sensors for real-time network traffic estimation and prediction: A sensitivity analysis. *Transportation Research Record*, 1981:253–259, 2006.

[14] NET. System wide adaptive ramp metering - high level design. Technical Report Final Draft, Prepared by NET for Caltrans and FHWA, 1996.

[15] E. Smaragdis and M. Papageorgiou. Series of new local ramp metering strategies. *Journal of the Transportation Research Board*, 1856:74–86, 2003.

[16] L. Chu, X. Liu, W. Recker, and H.M. Zhang. Performance evaluating of adaptive ramp metering algorithms using microscopic traffic simulation model. *Journal of Transportation Engineering*, 130(3):330–338, 2004.

[17] S. Ahn, R.L. Bertini, B. Auffray, J.H. Ross, and O. Eshel. Evaluating the benefits of a system-wide adaptive ramp-metering strategy in portland. *Journal of Transportation Research Board: Transportation Research Record*, 2012:47–56, 2007.

[18] S. Amin et al. Mobile century-using gps mobile phones as traffic sensors: a field experiment. In *Proceedings of the 15th World congress on ITS*, 2008.

[19] D. Work, O.P. Tossavainen, S. Blandin, A. Bayen, T. Iwuchukwu, and K. Tracton. An ensemble kalman filtering approach to highway traffic estimation using gps enabled mobile devices. In *Proceedings of the 47th IEEE Conference on Decision and Control*, 2008.

[20] http://pems.eecs.berkeley.edu/public/.