OPTIMAL SENSOR LOCATIONS
FOR ADVANCED TRUCK SURVEILLANCE ON CALIFORNIA FREEWAYS

Jaeyoung Jung, Ph.D.
Andre Tok, Ph.D.
Stephen G. Ritchie, Ph.D.

Institute of Transportation Studies
University of California, Irvine; Irvine, CA 92697-3600, U.S.A.
jaeyounj@uci.edu, ytok@uci.edu, sritchie@uci.edu

August, 2013

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu
OPTIMAL SENSOR LOCATIONS
FOR ADVANCED TRUCK SURVEILLANCE ON CALIFORNIA FREEWAYS

Jaeyoung Jung
Institute of Transportation Studies
University of California
Irvine, CA 92697-3600
Phone: +1–949–824–9806
FAX: +1–949–824–8385
Email: jaeyounj@uci.edu

Andre Tok
Institute of Transportation Studies
University of California
Irvine, CA 92697-3600
Phone: +1–949–824–9806
FAX: +1–949–824–8385
Email: ytok@uci.edu

Stephen G. Ritchie
Institute of Transportation Studies and
Department of Civil and Environmental Engineering
University of California
Irvine, CA 92697-3600
Phone: +1–949–824–4214
FAX: +1–949–824–8385
Email: sritchie@uci.edu

Paper length: 5,590 + 10*250 (figures and tables) = 8,090 words

Submitted for publication and presentation at the Transportation Research Board 93rd Annual Meeting
and Transportation Research Record
ABSTRACT

A new hybrid sensor technology integrating existing Weigh-In-Motion (WIM) axle configuration data with inductive signature data obtained from advanced Inductive Loop Detector (ILD) is gaining interest due to its potential to provide detailed classification of truck body types as well as anonymous tracking of truck movements on freeways. This paper describes the methodologies and analysis of two alternative strategies for optimal deployment locations for this new technology at existing WIM locations by utilizing sampled truck GPS trajectories on California freeways: (1) Flow-interception approach to maximize the total amount of net origin-destination (OD) flows captured; (2) Re-identification approach to maximize insights into origins and destinations of sampled truck trips, as well as routes of those trips.

The truck GPS samples used in this study is obtained from the American Transportation Research Institute (ATRI), which provides position and time stamp information of truck movements. The model designed for flow-interception is capable of selecting locations emphasizing different body types by employing the flow-based weight factor. The RSP model investigates the best locations for heavy truck movement identification on freeways by selecting pairwise locations, and is shown to be sensitive to the re-identification decay factor assumed.
Optimal Sensor Locations for Advanced Truck Surveillance on California Freeways

1. INTRODUCTION

Commercial vehicles represent a small fraction of vehicular traffic on most roadways. However, their influence on the economy, environment, traffic performance, and infrastructure are much more significant than their numerical presence suggests. Freight movements increasingly contribute air pollution. There has been increased concern about health and environmental impacts due to emissions of NO\textsubscript{x} and PM from heavy commercial vehicles, which are predominantly powered by diesel engines. Also commercial vehicles are the mode of choice for freight transportation in California: The largest amount of truck freight shipments in the U.S. by value, with $626 billion (outbound) and $619 billion (inbound), and by weight, with 768 million tons (outbound) and 782 million tons (inbound) (USDOT, 2003), and their traffic volume is expected to grow throughout the state over the next 20 years (FHWA, 2002). Hence, the movement of the commercial vehicle transportation system has direct implication on the development of advanced sensor technology, especially in conjunction with WIM (Weight-in-Motion) stations.

Truck travel data is critical across a wide spectrum of application needs, such as freight forecasting, regional transportation planning, air quality assessment, and infrastructure management. However, the availability of such data is regrettably restricted to axle-based classification counts and weight measures at Weigh-In-Motion (WIM) stations, together with estimates derived from factored volume measures obtained from inductive loop detectors (ILD) and automated vehicle classifiers. Although truck GPS data is available as well, and provides an excellent source for determining truck travel times along major corridors, they only represent a sample of trucks that participate in location reporting programs. Hence, the data is inadequate for providing a full representation of truck activity at any location or region. In addition, privacy concerns regarding truck participation require that the data be scrubbed to maintain anonymity. As a consequence, information that can be used to infer trip purpose, such as truck body configuration is not represented and cannot be inferred. Since present classification data from WIM stations is limited to axle-based configuration and gross axle weights, they cannot provide direct information associated with truck body configuration. However, truck body configuration is essential for determining and understanding truck travel behavior. The association of truck travel characteristics (such as vehicle miles traveled, payloads) with body configuration has been affirmed by Vehicle Inventory and Use Survey (VIUS) data. But, although VIUS provides arguably the most relevant and comprehensive data for understanding the impacts and implications of trucks by body configuration detail, the most recent effort is already over a decade old. In addition, the nature of the dataset – being derived from surveys by truck carriers – does not allow the assessment of specific corridors or regions for detailed truck impacts.

A new hybrid sensor technology is gaining interest due to its potential to provide detailed classification of truck body configurations and anonymous tracking of trucks over long distances without requiring in-vehicle instrumentation\textsuperscript{12}. This technology integrates existing WIM axle configuration data from WIM controllers with inductive signature data obtained from advanced signature capable inductive loop signature cards and has shown promising preliminary results in advanced truck body classification. In addition, re-identification of trucks using WIM data alone has been demonstrated over long distances across WIM stations located as much as 145 miles apart, and is expected to improve when enhanced with inductive signature data (1).

\textsuperscript{1} http://www.arb.ca.gov/board/books/2012/012612/prores1207.pdf
\textsuperscript{2} http://www.volpe.dot.gov/sbir/sol12_2/topics.html#FH4
This study presents two alternative methods based on the analysis of sampled truck GPS data to choose optimal deployment locations for this new technology at existing WIM locations, reflecting on two different deployment strategies that may be of interest to state agencies. In the first method, called the flow-interception approach, the objective is to deploy a selected number of locations that will maximize the total number of trip samples captured. Such an implementation would be desirable for an agency that would like to survey the highest possible representation of trucks traveling on the network. The second method—the re-identification approach—attempts to locate the detector stations such that deployed locations have the potential to capture not only the maximum number of trip segments, but also have the captured segments reflect the largest possible portion of each trip as well, constrained to the locations available for deployment. Decay functions were estimated to account for the effect of re-identification performance decay over large distances, since this is not yet available in literature. This implementation strategy is designed to maximize insights into origins and destinations, as well as routes of truck trips in the network.

2. TRUCK GPS DATA

The heavy truck GPS data used in this study was obtained from the American Transportation Research Institute (ATRI). In this dataset, anonymous randomly generated identification numbers (IDs) are used to maintain the confidentiality of truckers and trucking companies. Position and timestamp information is received from trucks equipped with automatic vehicle location equipment at predetermined intervals, which vary by truck. This dataset comprises of four weeklong subsets in the middle of each quarter in 2010 to provide seasonal representation of truck travel activities within the State of California, yielding a total of over 8 million truck positions. It is assumed that the truck trajectories derived from this dataset provide a reasonable representation of heavy truck activity in the State of California. The truck trajectories were subsequently matched with WIM locations to determine the number of trips captured by each WIM station, as well as the ordered sequence of WIM stations traversed by each truck trip for the development of the models described in this paper. Figure 1 shows the truck trajectories and locations of existing WIM stations in the State of California.
3. SENSOR LOCATION PROBLEM

Sensor placement problems have been extensively studied in transportation research in recent years as sensor technologies (e.g., Loop detectors, Image detectors, RFID, WIM) have been actively deployed on road networks. The location allocation of roadway sensors is a critical issue when limited budget restricts the number sensors that can be deployed.

Traffic origin-destination (OD) flow estimation is one of the important issues that emphasize the importance of sensor placement problems. The objective of the sensor placement problem for ODE is to maximize the measurement coverage of the passenger flows on a route segment with deployed sensor technologies in (2, 3). Yang et al. (4) and Chen et al. (5) studied traffic counting location problem to improve the OD trip table estimation method based on screen-line-based models. For traffic surveillance based on the concept of paired sensors, Li and Ouyang in (6) investigated optimal sensor locations for travel time estimation by maximizing sensor coverage and Liu and Danczyk in (7) studied a sensor allocation problem for freeway bottleneck identification. Gentili and Mirchandani introduced a new set of network sensor location problems to monitor particular classes of traffic flows in (8) and conducted a comprehensive study for sensor deployment mainly focused on network flow-observation and flow-estimation in (9). These location problems and the associated models in the literature have great potential to derive variations for different purpose of transportation research. Depending on the type of applications, different objective and constraints can be formulated.
Flow-interception Sensor Placement (FSP)

Flow-interception Sensor Placement (FSP) in this study is designed to maximize the total amount of net OD trips captured on a network. Genetili and Mirchandani (9) reviewed a series of sensor location problems and categorized those problems into two main types; (1) Full Flow-Observability problem and (2) Partial Flow-Observability problem, where the former minimizes the number of sensors to cover the entire network flow level and the latter tries to maximize intercepting flow volumes given the number of sensors. In (9), two Partial Flow-Observability problems (M3 and M6) are considered for FSP application. However, M6 is found to be a greedy strategy selecting locations without considering link flow dependence, resulting in double-counting trips that traverse more than one sensor. So, M3 is used as the base model for the flow-interception sensor placement (FSP) where sensors are located on links such that maximal net route flows are measured together with a positive fraction of trips between any observed OD pairs. Let \( l \) and \( j \) denote the set of all travel paths with non-zero flows and the set of candidate sensor locations on the network, respectively. Next, let \( f_i \) represent the traffic volume along path \( i \in l \). The FSP formulation is as follows:

\[
\begin{align*}
\text{Maximize} & \quad \sum_{i \in l} \mu_i y_i \\
\text{s.t.} & \quad \sum_{j \in j_x} x_j \geq y_i \quad \forall i \in l \\
& \quad \sum_{j \in j} x_j = P \\
& \quad \sum_{j \in j_x} \delta^j_i x_j \geq 1 \quad \forall i \in l_r \\
& \quad y_i \in \{0, 1\} \quad \forall i \in l \\
& \quad x_j \in \{0, 1\} \quad \forall j \in j \\
& \quad \delta^j_i \in \{0, 1\} \quad \forall i \in l, \forall j \in j \\
\end{align*}
\]

where \( \mu_i \) is the flow-based weight factor for different vehicle types on path \( i \). The objective function in Equation (1) maximizes the total intercepted net path flow for all given OD pairs. Two binary decision
variables, \( x_j \) and \( y_i \) are included, which determine whether the candidate location \( j \) is selected and whether path \( i \) is covered by sensor placement, respectively. Constraint (2-a) ensures that the decision variable \( y_i \) to be equal to zero if there is no sensor placed on path \( i \). Constraint (2-b) locates exactly \( P \) sensors on links while constraint (2-c) forces a particular set of paths \( I_r \) to be covered if needed. The incident parameter, \( \delta_{ij} \), indicates one if the candidate location \( j \) is on path \( i \), and zero otherwise. It is noted that constraint (2-c) can conflict with (2-b) if \( P \) is not large enough to cover \( I_r \) \((P < |I_r|)\). A set of vehicle types, \( B \) denotes possible types of vehicles passing over the candidate sensor locations. A proportion of flow for vehicle type \( b \in B \) on path \( i \) can be expressed as \( \rho_{ib} \). The flow-based weight factor, \( \mu_i \) can be defined as a flow multiplied by weight factors for each proportion of vehicle types. If the weight factor, \( w_b \) is the same for all vehicle types, then the model becomes an ordinary flow-interception model considering all vehicle types equally. It is noted that without constraint (2-c), the problem represents the maximum covering location problem, which is known as NP-hard.

**Heavy Truck Body Type Distribution**

One of the main requirements of FSP is the truck type distribution associated with each vehicle path. However, the GPS truck trajectories do not provide information on vehicle body types and associated commodity types. Hence, the body type information was obtained from an ongoing effort to develop the California Statewide Freight Forecasting Model\(^3\) (CSFFM). CSFFM is a commodity based model to forecast and analyze freight movement in and out of California. The model outputs include location-based commodity productions, consumptions, commodity flows, and vehicle flows by different transportation mode (truck, rail, and air). To identify trucks’ OD and paths, the truck position records in truck GPS trajectories were then associated with the CSFFM’s freight analysis zones (97 FAZ polygons in California) based on their GPS points. In addition to OD match, each truck trajectory is projected on CSFFM’s statewide transportation network to identify an ordered set of WIM stations are associated with each truck route.

CSFFM payload factors and body types were employed to determine the fraction of body types associated with each OD pair. The commodity-based payload factors were derived from the California portion of the 2002 Vehicle Inventory Use Survey (VIUS), which provide conversion factors from the total tonnage value associated with each of fifteen commodity groups defined in CSFFM to the corresponding number of vehicles for each OD pair. Table 1 shows the eight different body types defined in CSFFM. The proportions of body types based on the California portion of VIUS are subsequently applied to CSFFM commodity groups in Table 2.

\(^3\) [http://www.dot.ca.gov/hq/tsip/otfa/csffm/index.html](http://www.dot.ca.gov/hq/tsip/otfa/csffm/index.html)
Table 1. VIUS Body Types in CSFFM

<table>
<thead>
<tr>
<th>Body Types</th>
<th>CA portion of VIUS</th>
<th>National VIUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Concrete mixer and concrete pumper</td>
<td>4.49%</td>
<td>3.62%</td>
</tr>
<tr>
<td>2 Dump</td>
<td>12.56%</td>
<td>21.27%</td>
</tr>
<tr>
<td>3 Flatbed, stake, platform, and low boy</td>
<td>27.30%</td>
<td>19.49%</td>
</tr>
<tr>
<td>4 Service</td>
<td>6.34%</td>
<td>4.52%</td>
</tr>
<tr>
<td>5 Tank</td>
<td>23.73%</td>
<td>8.28%</td>
</tr>
<tr>
<td>6 Trash, garbage, recycling, and vacuum</td>
<td>5.53%</td>
<td>4.21%</td>
</tr>
<tr>
<td>7 Van</td>
<td>17.05%</td>
<td>30.00%</td>
</tr>
<tr>
<td>8 Others</td>
<td>3.00%</td>
<td>8.61%</td>
</tr>
<tr>
<td>Total</td>
<td>100.00 %</td>
<td>100.00 %</td>
</tr>
</tbody>
</table>

Table 2. Body Type Distribution by Commodity Group (FHWA Class 9)

<table>
<thead>
<tr>
<th>Commodity Group</th>
<th>Body Type</th>
<th>%</th>
<th>Commodity Group</th>
<th>Body Type</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Agriculture products</td>
<td>2</td>
<td>0.0185</td>
<td>8</td>
<td>Manufactured Products</td>
<td>0.0787</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.2407</td>
<td>3</td>
<td>0.0787</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0926</td>
<td>7</td>
<td>0.4494</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.5741</td>
<td>8</td>
<td>0.1348</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.0741</td>
<td>9</td>
<td>0.0345</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 Non-metallic mineral</td>
<td>1.0000</td>
<td>11 Metal manufactured products</td>
<td>0.5517</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.7111</td>
<td>2</td>
<td>0.0519</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.1333</td>
<td>3</td>
<td>0.4545</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0444</td>
<td>7</td>
<td>0.6897</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.1111</td>
<td>10</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 Crude petroleum</td>
<td>1.0000</td>
<td>12 Waste materials</td>
<td>0.5517</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0185</td>
<td>2</td>
<td>0.5517</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.2857</td>
<td>3</td>
<td>0.4545</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.1786</td>
<td>7</td>
<td>0.1724</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 Fuel, Oil products</td>
<td>0.5000</td>
<td>8</td>
<td>0.0390</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.1333</td>
<td>7</td>
<td>0.2500</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0457</td>
<td>3</td>
<td>0.1724</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.8889</td>
<td>7</td>
<td>0.2500</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.0556</td>
<td>13 Electronics</td>
<td>0.6552</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 Gravel, Non-Metallic minerals</td>
<td>0.0556</td>
<td>5</td>
<td>0.1087</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.5556</td>
<td>7</td>
<td>0.8261</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.7222</td>
<td>14 Transportation equipment</td>
<td>0.5000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 Coal, Metallic minerals</td>
<td>0.1111</td>
<td>2</td>
<td>0.2500</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.1111</td>
<td>3</td>
<td>0.2500</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0556</td>
<td>7</td>
<td>0.5000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.2391</td>
<td>15</td>
<td>0.8571</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 Food, Beverage, Tobacco products</td>
<td>0.2174</td>
<td>5</td>
<td>0.1429</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.5435</td>
<td>7</td>
<td>0.5000</td>
<td></td>
</tr>
</tbody>
</table>

The following method is applied for the calculation of $\hat{\rho}_b^k$ stated in FSP:

$$T_b^k = \frac{m_{k1} \varphi_{1b}}{p_{1b}} + \frac{m_{k2} \varphi_{2b}}{p_{2b}} + \frac{m_{k3} \varphi_{3b}}{p_{3b}} + \cdots + \frac{m_{k15} \varphi_{15b}}{p_{15b}} = \sum_{s \in S} \frac{m_{ks} \varphi_{sb}}{p_{sb}}$$  \hspace{1cm} (3-a)

$$\rho_b^k = \frac{T_b^k}{T^k}$$  \hspace{1cm} (3-b)

$k$: $k$-th OD pair, $k \in K$

$s$: commodity group, $s \in S$

$T_b^k$: number of trucks in vehicle type $b$ for $k$-th OD pair
Let $T^k_b$ denote the number of trucks in truck type $b$ for the given OD pair $k$, $k \in K$, defined in (3-a), which can be calculated by the summation of all commodity groups. The estimated proportion of body type $b$ for OD pair $k$, $\rho^{b}_k$ can be obtained from Equation (3-b), where $T^k$ represents a total number of vehicles assumed for $k$. The same proportion is assumed for all possible paths between OD pair $k, i_k \in k$.

**Re-identification Sensor Placement (RSP)**

Vehicle re-identification is an emerging advanced sensor technology that aims to match vehicles crossing two different locations based on vehicle attribute data. ILD based vehicle re-identification methods using individual vehicle’s inductance signature over an ILD have shown great feasibility by (10, 11, 12, 13). The benefit of re-identification of vehicles is known for trip OD matrix estimation and vehicle path flow reconstruction (identification) on transportation network.

For truck re-identification, WIM has been known as a major method to identify heavy truck’s data. There has been interest in developing an inductive loop signature-WIM based heavy truck monitoring system, in which the conventional loop detectors and WIM stations are combined. Such an integrated technology can offer a great potential to identify heavy truck movements on the freeway. Cetin et al. (1) studied the re-identification of trucks over long distance can be performed. They developed algorithms to match commercial vehicles that cross two WIM stations with a higher accuracy in Oregon by using vehicle length and axle information.

The objective function and constraints employed in RSP are based on the study by Sherali et al. in (14) and Liu and Danczyk in (7), where it is formulated as a non-linear optimization problem. The objective function maximizes the benefits that can be represented as the utility of the placement of two sensors. In Sherali’s study, benefit factors are represented to capture travel time variability on the transportation network while Liu and Danczyk used a benefit factor to measure the non-negative speed gradient on a highway stretch in order to identify freeway bottlenecks.

Since the main purpose of re-identification is to identify vehicle movements on freeways, the objective of the RSP problem is to select optimal pairwise locations. First, we assume that the model mainly considers a benefit value in terms of capturing as many truck routes as possible given that the truck routes known from ground truth sampling. Also, constraints reflect key factors affecting the accuracy of re-identification, such as distance and traffic flow.

Let $l$ be the set of OD paths on the network. Each path $i \in l$ is specified by its traffic volume $f_i$, which is assumed to be known via GPS trajectory sampling. If $f_i$ passes at least two sensors, the route portion between two sensors can be identified as a vehicle re-identification route. Let $J_i$ denote the set of all candidate locations for re-identification sensors. Each path $i$ passes a set of candidate sensors on highway network, $J_i$. The RSP formulation is following:

---

4 [http://www.volpe.dot.gov/sbir/sol12_2/topics.html#FH4](http://www.volpe.dot.gov/sbir/sol12_2/topics.html#FH4)
Jung, Tok, and Ritchie

RSP-1) Maximize $\sum_{i \in I} \sum_{h \in J_i} \sum_{e \in J_i} b_{he} x_h x_e$ 

\[ \text{s.t.} \quad \sum_{j \in J} x_j \leq P \]
\[ \sum_{j \in J} c_j x_j \leq U \]
\[ x_j \in \{0,1\} \quad \forall j \in J \]

$b_{he}$ = benefit factor between head ($h$) and rear ($e$) sensors

$P$ = maximum number of sensors that can be installed on freeways

$c_j$ = unit cost of installing sensors at any site $j$

$U$ = maximum installing budget

We assign the upstream and downstream sensor traversed by a truck trip as the head ($h \in J_i$) and rear ($e \in J_i$) sensors, respectively. If there exist at least two sensors at $x_h$ and $x_e$ along path $i$, its utility can be recorded as $b_{he}$. The objective function (4) seeks to maximize the total benefit based on sensor placements among candidate sites, $J$. Constraint (5-a) asserts that the total number of sensors should not exceed the available maximum number, $R$, while constraint (5-b) limits the total cost of installing sensors at each site $j$ with a unit cost $c_j$. The decision variable $x_j$ is defined as binary, where $1$ represents the placement of a sensor at location $j$ and $0$ otherwise in (5-c).

The benefit factor ($b_{he}$) constituting the objective function parameters here reflects the ability to capture highway paths regarding the variability of freight truck movement. Since the number of sensors on path $i$ is a key factor in estimating the true OD paths, the locations of upstream and downstream sensors are critical. A path coverage concept is introduced as a benefit factor of RSP. Consider that the length of an individual vehicle path $i$ connecting its true origin and destination is denoted by $L_i$. Then, let $R_i$ denote a set of candidate sensors on path $i$. These sensors play a critical role in reconstructing the individual path.

If there are two sensors placed on path $i$, denoting $R_i = (r_h, r_e)$ for upstream (head) and downstream (rear), the identified distance by two sensors, $d_i(r_h, r_e)$ along path $i$ can be calculated by actual vehicles trajectories along truck routes. If $d_i(r_h, r_e)$ represents a significant proportion of $L_i$, the benefit will be more appealing for the path identification as described in (6). On the other hand, if $d_i(r_h, r_e)$ is very short compared to $L_i$, it can be inferred that the two sensors do not provide much information about path $i$.

$$b_{he}^d = \begin{cases} 0, & r_e < r_h \\ \frac{d_i(r_h, r_e)}{L_i}, & \text{otherwise} \end{cases}$$

There are currently no comprehensive empirical increase for re-identification performance combined with WIM and ILD signature. However, a recent study by Cetin (15) evaluated factors which affect the matching accuracy of truck between WIM station pairs, such as distance, travel time variability, truck volumes, and sensor accuracy and consistency. They demonstrated that re-identification with WIM can be performed over large distances, and WIM sensor accuracy and traffic volume were identified as the major factors determining the search space at the upstream location.

In this study, the variability of travel time between upstream and downstream is considered to potentially affect the accuracy of re-identification for WIM and ILD signature technology because a search space for
a downstream vehicle is determined based on the travel time stamps at head and rear stations. The width of time windows usually increases with the distance between upstream and downstream detector increases, generally yielding a lower probability of vehicle matching. Therefore, one assumption is that the error rate of re-identification increases as the distance between head and rear sensors increase in (7).

\[
\text{Re-identification Error Rates } \propto d_i(r_h, r_e) \tag{7}
\]

\[
g_1(r_h, r_e) = e^{-s/\tau} \tag{8}
\]

We assume a distance-based exponential decay function shown in Equation (8), where \(\tau\) is a decay rate with a positive constant. Let \(s\) denote an input variable for the function, \(g\). In this study, we assume that \(s\) can be replaced with \(d_i(r_h, r_e)\). Figure 2(a) shows three performance decay curves assumed by different parameter sets with short (\(\tau=100\)), medium (\(\tau=200\)), and long distance performance (\(\tau=400\)).

A second approach assumes that the decay factor, \(g_2\), can be a location-specific traffic flow from upstream to downstream. Suppose that two detectors, \([r_h, r_e]\), are installed on a freeway. Two data types can be used to estimate the likelihood of finding the correct match—traffic flow from upstream (\(h\)) to downstream (\(e\)), \(f_{he}\) and all traffic counts at both upstream and downstream sites, \(f_{hej}\) and \(f_{ej}\) in Equation (9). The basic idea of this approach is simply assuming a search space given the number of vehicles coming from the upstream. For example, if two locations are close and the most vehicles passing the upstream location are going to the downstream location, there would be higher probability that the downstream vehicles can be found at the upstream candidate set, which means that both \(f_{he}\) and \(f_{hej}\) are higher. On the other hand, if the majority of volumes at the upstream site take other paths instead of going to the downstream site, there would be fewer downstream vehicles coming from the upstream sites, resulting in a higher \(f_{hej}\), but lower \(f_{he}\). The same concept can be extended to the traffic count (\(f_{ej}\)) at the downstream site. Note that \(g_2\) is a location-specific factor estimated from the ground truth samples. The observed traffic counts and the flow between upstream and downstream can be abstracted from truck trajectories.
Segment combinations of possible upstream and downstream WIM sites were extracted from all truck trajectories. A total of 151,760 segments were identified with 2,460 segments of unique upstream and downstream WIM site pairs. Figure 2(b) shows the assumed performance decay \( g_2 \) for each combination of upstream and downstream by distance. As expected, higher rates are generally found within the shorter distance range. The highest value is 0.309 corresponding to truck movement between site 30 (Mt. Shasta on I-5 South) and site 2 (Redding on I-5 South). Almost 60 percent of vehicles that passed the upstream site also traversed the downstream site. The performance decay associated with WIM site pairs is independent of the presence of other upstream or downstream sites.

So far, only two sensor examples are illustrated, but multiple sensors can be utilized for one path, \(|R_i| > 2\). For a subset of sensor pairing, we assume that there are \(|R_i|-1\) levels of pairs for benefit factors. For the example of \( R_i = \{r_1, r_2, r_3, \ldots, r_s\} \), there are \( s-1 \) benefit factors contributed sequentially by each sensor-based segment \( \{r_1, r_2\}, \{r_2, r_3\}, \ldots, \{r_{s-1}, r_s\} \). The total benefit for multiple sensors on a path can be summed up as described in Equation (10).

\[
b_{he} = \sum_{h,e \in R_i} b_{he}^d g(h, e)
\]  

(10)

4. NUMERICAL EXPERIMENTS

FSP Solution

The proposed FSP problem takes a linear programming form. This is useful for finding an exact solution using conventional linear programming solvers if the problem is not extremely large in scale. Our preliminary computational experience with C++ and CPLEX 12.4 Concert Library showed reasonable computational times around 5 minutes on Intel Core 2 Duo 2.6 GHz CPU and Windows XP. The origins and destinations of 131,201 truck trajectories are grouped by 97 CSFFM Freight Analysis Zones (FAZs) in California. A total of 13,009 unique OD paths were identified with 83 distinct WIM locations. Hence, each OD pair could be associated with multiple distinct paths. Truck trajectories with one end of their trip outside of California have the external end matched to the zones closest to the state boundary. The number of WIM sites traversed by paths range between one and eleven sites.

FSP Results for Heavy Truck Body Types

The aforementioned body type distribution in Table 2 was used to estimate the distribution of body types for each commodity group. Truck body types G1, G4, and G6 were not considered as they are not associated with freight movement. The selection priorities were performed using a stepwise approach to increase the maximum number of locations selected. Table 3(a) and (b) reports results for one through five and twenty selected locations, using six different criteria for selecting the optimal locations. In the first, all truck body types are given equal priorities for flow interception. The remaining criteria correspond to having only one of the body types corresponding to G2, G3, G5 G7 and G8, respectively, be desired in determining the optimal locations.
Table 3. Comparison of FSP location selection results for different truck body type priorities

(a) 1 to 5 selected stations

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Selected WIM Sites</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 2$</td>
<td>$P = 3$</td>
<td>$P = 4$</td>
<td>$P = 5$</td>
<td></td>
</tr>
<tr>
<td>All groups</td>
<td>95</td>
<td>95, 102</td>
<td>82, 95, 102</td>
<td>3, 82, 95, 102</td>
<td>3, 77, 82, 95, 102</td>
<td></td>
</tr>
<tr>
<td>G2 Dump</td>
<td>47</td>
<td>47, 102</td>
<td>47, 95, 102</td>
<td>44, 47, 95, 10</td>
<td>44, 47, 95, 102, 103</td>
<td></td>
</tr>
<tr>
<td>G3 Flatbed, stake, platform and lowboy</td>
<td>95</td>
<td>82, 95</td>
<td>82, 95, 102</td>
<td>44, 82, 95, 102</td>
<td>44, 77, 82, 95, 102</td>
<td></td>
</tr>
<tr>
<td>G5 Tank</td>
<td>95</td>
<td>82, 95</td>
<td>82, 95, 102</td>
<td>44, 82, 95, 102</td>
<td>44, 77, 82, 95, 102</td>
<td></td>
</tr>
<tr>
<td>G7 Van</td>
<td>95</td>
<td>3, 95</td>
<td>3, 77, 95</td>
<td>3, 77, 95, 102</td>
<td>3, 77, 82, 95, 102</td>
<td></td>
</tr>
<tr>
<td>G8 Others</td>
<td>82</td>
<td>82, 102</td>
<td>82, 95, 102</td>
<td>1, 82, 95, 102</td>
<td>3, 44, 82, 95, 102</td>
<td></td>
</tr>
</tbody>
</table>

* [95]: Ontario on CA SR-60, [47]: Castaic on I-5, [82]: Glendora on I-210, [102]: Delhi on CA SR-99, [3]: Antelope on I-80

(b) 20 selected stations

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Selected WIM Sites</th>
<th>$P = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All groups</td>
<td>3, 5, 10, 15, 24, 29, 37, 44, 47, 57, 59, 63, 67, 69, 75, 77, 82, 95, 102, 114</td>
<td></td>
</tr>
<tr>
<td>G2 Dump</td>
<td>3, 7, 10, 15, 17, 29, 44, 47, 55, 57, 59, 61, 68, 75, 82, 87, 95, 103, 111, 114</td>
<td></td>
</tr>
<tr>
<td>G3 Flatbed, stake, platform and lowboy</td>
<td>3, 5, 7, 10, 15, 24, 29, 37, 41, 44, 47, 59, 61, 67, 69, 75, 77, 82, 95, 103</td>
<td></td>
</tr>
<tr>
<td>G5 Tank</td>
<td>1, 5, 7, 10, 15, 24, 35, 37, 41, 44, 46, 47, 59, 61, 67, 69, 77, 82, 95, 102</td>
<td></td>
</tr>
<tr>
<td>G7 Van</td>
<td>3, 5, 10, 15, 24, 29, 30, 37, 44, 47, 57, 59, 63, 67, 69, 75, 77, 82, 95, 114</td>
<td></td>
</tr>
<tr>
<td>G8 Others</td>
<td>2, 3, 5, 7, 10, 15, 29, 35, 41, 44, 47, 59, 61, 67, 69, 77, 82, 95, 102, 114</td>
<td></td>
</tr>
</tbody>
</table>

The results show that site 95 (Ontario on CA SR-60) is ranked with the highest priority across the criteria corresponding to all groups, G3, G5, and G7. This indicates that the model is sensitive to different body type priorities. The location is also reported as one of the top three ranked sites (95, 77, and 82) of FHWA Class 9–5-axle tractors pulling semi-trailers – truck counts obtained from 2010 WIM data, indicating an agreement between the truck GPS data with WIM data. Although site 95 is not initially picked under criteria G2 and G8, it is selected in $P = 3$, indicating that it is generally an ideal site for flow interception. Also, site 102 (Delhi on CA SR-99) is selected across all criteria for $P = 5$, but is not included in $P=20$ for all but two criteria: G5 and G8. That indicates that a significant proportion of truck trips traversing site 102 may be captured by one or more combination of alternative sensors in close proximity. Indeed, site 75 (Keyes on CA SR-99), which is located within 13 miles north of 102 on SR-99 in Central California, is selected for the other criteria in the absence of site 102, and both these sites are never chosen simultaneously for each criterion. This confirms that the FSP model gives priority to selecting non-overlapped locations.

Figure 3 shows a comparison of site selection results using the six different selection criteria for selecting five optimal locations. Sites chosen by one or more of the selection criteria have their corresponding slices shaded white. Both sites 95 and 102 are chosen across all criteria, while site 47 (Castaic on I-5) and 103 (Orange on CA SR-57) are chosen only when optimizing the location selection for G8 (Other) trucks.
A sensitivity analysis was performed across a range of sensor deployments, for the number of selected sites, $P$ ranging from 1 to 40. Figure 4 shows the flow-interception and the path-interception using the six different criteria. When deploying the body type classification technology for up to 40 locations, the flow-interception can capture more than 90 percent of the given truck flows. G7 shows the dominant flows among the considered body types. When the G7 criterion is used and deployed at 40 locations, the G7 trucks captured by these selected locations represent 54 percent of total flows. There is no significant difference in path interception across all criteria.
RSP Solution

The RSP problem is a constrained mixed-integer quadratic programming problem. Since the benefit factors in RSP can vary depending on the combination of selected locations, the complexity of benefit makes it a challenge to solve with traditional nonlinear solvers. Genetic Algorithm (GA) is an adaptive meta-heuristic algorithm used to solve the proposed combinatorial optimization problem in (16). The GA method involves a chromosome structure that represents a solution of the problem and the solution evolves by copying chromosomes and swapping partial chromosomes over generations. Although GAs do not guarantee optimal solutions, they are efficient in seeking approximate solutions in combinatorial optimization problems. Each chromosome is made up of candidate locations representing binary decision variables. An initial population is given by randomly distributed servers over the candidate locations. For genetic operators, Elitism, Crossover and Mutation techniques are employed with the maximum number of generations (e.g., 500 generation span) for each scenario.

RSP Results

As shown in Table 4, three types of re-identification performance decay scenarios were considered, with \( g_0, g_1, \) and \( g_2 \) representing no decay, distance-based decay, trip-based location specific decay, respectively. Unlike FSP, only trips traversing two or more WIM stations on an observed path were considered. A total of 58,762 individual trips were identified to have two or more sensors among 131,201 truck paths. The average number of WIM stations on a path is 2.62 stations. 35,012 (59.6%), 15,252 (26.0%), and 5,898 (10.0%) trajectories contained two, three, and four WIM sites, respectively, while only about 2,600 (4.4%) trajectories contained five or more WIM sites.
Table 4. RSP location selection results

<table>
<thead>
<tr>
<th>Number of</th>
<th>( P )</th>
<th>( g_{0} ) (No decay)</th>
<th>( g_{0} ), ( \tau = 400 )</th>
<th>( g_{0} ), ( \tau = 200 )</th>
<th>( g_{0} ), ( \tau = 100 )</th>
<th>( g_{0} ) (Trip-based)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selections</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Southern California Locations
Table 4 reports the RSP location selection results showing between two and ten locations with different performance decay scenarios. New locations from the increase in selected locations are shown in red. For all scenarios, sites 41 and 57 are the best candidates for the maximum re-identification benefit if only two locations are allowed within budget. These two detectors are located 58 km apart along I-80 in Northern California. No difference can be found across scenarios in $P = [2, 3, 4]$. For $P = 5$, $g_1$ ($\tau=100$) and $g_2$ show different results with sites 51 ($g_1$) and 46 ($g_2$), although all selected sites are in Northern California. The solution generally adds an additional site based on the previous solutions, due to the pairwise selection characteristics. There are exceptions, however. For example, $g_2$ drops sites 46 and 72 from $P = 5$ and adds sites 10, 75, and 102 for $P = 6$. This reveals that the location selection can significantly change depending on the benefit from different pairwise combinations. There is also a preference for Northern California locations across all scenarios, with only $g_0$ and $g_2$ choosing Southern California locations starting at $P = 8$.

Figure 5 shows a spatial comparison the RSP location selection results obtained with $g_0$, $g_1$ ($\tau=100$), and $g_2$ for $P=10$. It shows that $g_0$ expands the re-identification network from San Francisco (site 57) to Los Angeles (site 82), whereas the result of $g_1$ remains in Northern California. For $g_2$, the location selection that started from sites 41 and 57 extends to site 47 in Northern Los Angeles when considering nine locations in Table 3(b). The three locations selected in Northern Los Angeles, 47, 73, and 74 are critical points for identifying the truck route diversion on I-5 and CA-99, which is an intuitive result.

Figure 5. Comparison of RSP location selection results for $g_0$, $g_1$, $\tau = 100$ and $g_2$, with ($P = 10$)
Figure 6 shows sensitivity analysis with up to 40 sites selection. First, it is noticed that the objective values in Figure 6(a) are not as smooth as the curves in Figure 4(a). That can be expected as the pairwise selection method is based on a meta-heuristic algorithm—GA. However, the results are very consistent with applied decay assumptions. Without the performance decay assumption ($g_0$), the result shows the highest objective values, and then $\tau = 400, 200, \text{ and } 100$ are ranked sequentially. As expected, $g_2$ shows the lowest objective values due to the trip-based performance decay assumption that considers a potential search space for candidate vehicles, but it is conceivable that $g_2$ is the most reliable because the assumption is based on the observed truck flows. Figure 6(b) shows path-length coverage (%) which can be defined as the distance of covered segments between upstream and downstream locations divided by the total segment length. The scenario $g_0$ shows covering more than 80 percent when 40 locations are installed. Not much difference in the path coverage can be seen because performance decay impacts are not taken into account in this figure, but the scenario with $\tau = 100$ shows the lowest values.

![Figure 6. RSP Results with Decay Performance Assumptions](image)

5. CONCLUSION

This paper addresses two applications for optimally allocating the hybrid sensor technologies integrating existing WIM with inductive loop signature data on California freeways: flow-interception sensor placement (FSP) and re-identification sensor placement (RSP). Truck GPS trajectories were utilized with 97 CSFFM FAZs to identify truck’s OD and paths. A total of 83 existing WIM stations were considered as candidate sites.

The FSP model is capable of selecting locations emphasizing different body types by employing the flow-based weight factor. We assumed the distribution for eight different truck body types based on 2002 VIUS data for the State of California.

The RSP model investigates the best locations for heavy truck movement identification on freeways by selecting pairwise locations, and is shown to be sensitive to the re-identification decay factor assumed. Individual truck trajectories were used as the ground truth for truck movement. We identified that sites 41 and 57 have a strong influence on the site selection including many candidate sites in Northern...
California. On the other hand, candidate locations in Southern California are not much attractive to identify truck movement given limited numbers of sensors even though higher traffic counts are measured. This may be because few of the existing WIM stations in Southern California are located sequentially. As a result, most combination pairs do not capture significant through flows despite high volumes at individual locations. The re-identification decay factors assumed in the RSP may not be representative of true re-identification performance. However, they can be easily substituted when such information becomes available through future studies.

Although the methods described in this paper were designed for the implementation of inductive signature technology at existing WIM stations, the concepts presented for FSP can be extended to the deployment of other truck surveillance infrastructure with the use of available truck GPS data. Similarly, the RSP can be applied to determine the optimal locations for deploying other truck tracking technologies such as Bluetooth or RFID. In these cases, re-identification decay would not be a concern.

ACKNOWLEDGEMENT

The research reported in this paper was supported by the California Air Resources Board (CARB). The authors gratefully acknowledge the assistance provided by Joseph Chow of Ryerson University and Nesamani Kalandiyur of CARB. The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of CARB. This paper does not constitute a standard, specification, or regulation.

REFERENCE


