Optimal Heterogeneous Sensor Deployment Strategy for Dynamic Origin-Destination Demand Estimation

Senlai Zhu, Ph.D. Candidate
School of Transportation, Southeast University
Si Pai Lou #2, Nanjing, China, 210096
Tel: +86-15850602236
Fax: +86-025-83795649
Email: zhusenlai@163.com

Hong Zheng, Ph.D.
Lyles School of Civil Engineering/NEXTRANS Center
3000 Kent Avenue
West Lafayette, IN 47906
Tel: +1-765-496-9768
Email: zheng225@purdue.edu

Srinivas Peeta, Ph.D., Professor (Corresponding Author)
Lyles School of Civil Engineering
550 Stadium Mall Drive
West Lafayette, IN 47907
Tel: +1-765-494-2209
Fax: +1-765-496-7996
Email: peeta@purdue.edu

Yuntao Guo, Ph.D. Candidate
Lyles School of Civil Engineering/NEXTRANS Center
3000 Kent Avenue
West Lafayette, IN 47906
Tel: +1-765-496-9768
Email: guo187@purdue.edu

Lin Cheng, Ph.D., Professor
School of Transportation, Southeast University
Si Pai Lou #2, Nanjing, China, 210096
Tel: +86-13951716936
Fax: +86-025-83795649
Email: gist@seu.edu.cn

Wei Sun, Ph.D., Associate Professor
School of Information and Control,
Nanjing University of Information Science & Technology
Ning Liu Rd. #219, Nanjing, China, 210044
Tel: +86-13739182534
E-mail: sunw0125@163.com

Submitted for Presentation and Publication consideration.
Revised, October 2015
Total number of words= 5983(text) + 1500 (4 figures, 2 tables) = 7,483 words
This paper proposes a network sensor location problem (NSLP) model to determine the optimal heterogeneous sensor deployment strategy in terms of the numbers of link (counting) and node (video/image) sensors and their installation locations for the dynamic origin-destination (O-D) demand estimates under a budget constraint. The proposed NSLP model also captures the impact of the time duration for which traffic data measurements are available on the optimal sensor deployment strategy, which has not been addressed in previous studies. A sequential sensor location algorithm that avoids matrix inversions is proposed to solve the NSLP model under the assumption of multivariate normal distribution for the prior dynamic O-D demand estimates. A network from a part of Chennai, India is used to illustrate the performance of the proposed NSLP model and solution algorithm. The results show that the proposed algorithm can identify the sensor locations based on the amount of information provided, and determine the numbers of each sensor type to maximize O-D demand estimation accuracy. They also illustrate that the proposed NSLP model can capture the time-varying characteristics of the amount of information provided by each sensor location. The proposed method can be used to analyze the impacts of various sensor types, their numbers, locations, and the time duration for which traffic data measurements are available for dynamic O-D demand estimation under limited resources.

Key words: network sensor location problem; heterogeneous sensor deployment strategy; dynamic O-D demand estimation; sequential sensor location algorithm
INTRODUCTION

Origin-destination (O-D) demand serves as an essential input for traffic assignment to predict the network traffic flow pattern. Reliable real-time O-D demand is useful for a range of traffic operational and management applications, such as real-time route guidance, and the online evaluation of various Intelligent Transportation System (ITS) strategies. Due to recent advances in real-time sensing technologies, different types of traffic data can be collected using a variety of sensors, such as GPS, Bluetooth, video camera, automatic vehicle identification (AVI), plate scanning, etc. With these sensing technologies, one can process the data to obtain traffic data measurements on links, turning movement counts at nodes, and full or partial vehicle trajectories along the path.

Based on the traffic flow characteristics of data collected from sensors, commonly used sensors can be categorized as follows. Counting sensors can count vehicles on a lane or a set of lanes in the network. This sensor type includes inductive loop detectors, magnetic detectors, etc. One can use vehicle count data to monitor network performance such as speed, density, occupancy, and flow rates. Image/video sensors can take images or videos of moving flows. For example, a fixed camera or video can be used to measure flows at an intersection. By processing the images or videos, it is possible to recognize moving vehicles in the scene. Vehicle-ID sensors can be used to identify vehicle IDs in the network. For instance, license plate readers (1) that use camera images, or Automatic Vehicle Identification (AVI) (2) readers that use radio-frequency identification (RFID) tags or bar-codes, can be located over lanes or on the roadside to identify vehicles. Based on the identified vehicles, full or partial path information such as path flows or travel times can be derived.

O-D demand has been estimated using traffic data measurements which can be obtained through sensor installation (2, 3, 4). However, in real-world applications, sensors cannot be installed on every link, node and/or path in the network due to limited budget. This motivates the need to optimally determine the locations of sensors in the network. As the locations of sensors in the network can significantly affect the accuracy and reliability of the O-D demand estimates, the network sensor location problem (NSLP) has received a lot of attention in recent years.

Most existing NSLP models are designed for static O-D demand estimation, and sensors are usually located on links only. Because the actual O-D demand is usually unknown in determining the sensor locations, indirect quality measures that do not need knowledge of the exact O-D demand are used in most existing NSLP models. For example, Lam and Lo (5) proposed to use traffic flow volume and O-D coverage criteria to determine priorities for locating sensors. Yang et al. (6) introduced the maximum possible relative error (MPRE) criterion to calculate the most possible deviation of the estimated O-D demand from the unknown true O-D demand. Yang and Zhou (7) further proposed four basic location rules, namely the maximal flow fraction rule, the O-D covering rule, the maximal flow interception rule, and the link independence rule. Yim and Lam (8) evaluated several of these rules on a large-scale network. Bianco et al. (9) proposed an iterative two-stage procedure and several priority-based greedy heuristics to cover the O-D demand and reduce the MPRE value. Gan et al. (10) introduced a modified MPRE formulation, termed the expected relative error (ERE), to represent the expected error between the true and estimated O-D demands. Bierlaire (11) proposed the total demand scale (TDS) measure to calculate the difference
between the maximum and minimum possible total demand estimates, which can be used for both static and dynamic O-D demand estimation. Chen et al. (12) extended the TDS measure to consider the quality measure at different spatial levels, and labeled it the generalized demand scale (GDS) measure. Simonelli et al. (13) proposed a synthetic dispersion measure (SDM), which is related to the trace of the covariance matrix of the posterior demand estimates conditional upon a set of sensor locations. Based on the trace of the covariance matrix of the posterior traffic flow estimates, Zhu et al. (14) proposed a stepwise method to identify sensor locations for the traffic flow estimation, including the O-D demand and the unobserved link flows.

Estimating time-dependent (dynamic) O-D demand is substantially more complicated compared to the static O-D estimation problem due to the time-varying characteristic. Hence, the corresponding NSLP is also difficult to address. Eisenman et al. (15) proposed a conceptual framework for the sensor location problem to minimize the error in the real-time O-D demand estimates. Fei et al. (16) extended Eisenman et al.’s (15) approach to examine the NSLP under two different scenarios (with and without budget constraints). The TDS measure proposed by Bierlaire (11) to locate sensors for the static O-D estimation problem was also used in their study to estimate the time-dependent O-D demand.

The NSLP has been investigated to handle different types of measurements such as AVI readers and license plate recognition techniques, mainly for the static O-D estimation problem. For instance, Chen et al. (17) proposed a multi-objective model for locating AVI readers on the network, which was extended by Chen et al. (18) to accommodate different travel demand patterns. Minguez et al. (19) sought to optimize the traffic plate scanning locations for O-D demand and route flow estimation under budget constraints. The traffic plate scanning location problem was also studied by Castillo et al. (1). Yang et al. (20) and Chen et al. (21) proposed addressing the screen line-based traffic counting location problem. Hu et al. (22) proposed a bi-level optimization model to solve the NSLP and determine an optimal deployment strategy for heterogeneous sensors (vehicle detector sensors and license plate recognition). For the dynamic O-D demand estimation problem, Asakura et al. (23) provided an off-line least-squares model to simultaneously determine the O-D demand and the identification rates of AVI data based on the locations of the AVI readers. Zhou and List (24) proposed a model for locating a limited set of traffic counting stations and AVI readers in a network so as to maximize the expected information gain for the dynamic O-D demand estimation problem.

In addition to its use in the traditional O-D demand estimation problem, NSLP has been used in other related domains. For example, Viti et al. (25) solved the sensor location problem for the travel time estimation problem. Xing et al. (26) proposed an information-theoretic sensor location model to minimize total travel time uncertainty. Bianco et al. (27) applied a genetic algorithm approach to identify sensor locations for estimating all link flows in the network. Castillo et al. (28) and Viti et al. (29) discussed the observability problem, to identify the set of sensor locations that would enable full O-D demand observability. He (30) proposed a graphical approach to locate sensors for link flow inference. Hu et al. (31) suggested a procedure that does not require any prior O-D demand matrix but entails explicit path enumeration, for the identification of all link flows using traffic data measurements on an optimal subset of links under a budget constraint. Ng (32) relaxed the
assumption of explicit path enumeration by using a node-based formulation for the Hu et al. (31) problem context.

In the literature, most NSLP models are restricted to one type of sensor, especially for the dynamic O-D demand estimation problem, where only traffic data measurements collected by link sensors has been considered. In addition, past studies have not considered the impact on the optimal sensor deployment strategy due to the time duration for which traffic data measurements are available. For example, past studies require traffic data measurements for a link for the entire time period of interest (such as the peak period, which can be in the order of hours) related to determining the dynamic O-D demand to identify the optimal sensor deployment strategy. However, in practice, traffic data measurements may be available for that link for only a much shorter time period (for example, one hour); the NSLP models proposed in past studies are restrictive in that they cannot be used to determine the optimal sensor deployment strategy without traffic data measurements for the entire time period. The optimal sensor deployment strategy may change when the time duration for which traffic data measurements are available is limited.

To address the aforementioned gaps, this paper proposes a NSLP model to identify the optimal heterogeneous sensor deployment strategy to maximize the quality of dynamic O-D demand estimates, or to minimize the variability of dynamic O-D demand estimates, under a limited budget. The sensor deployment strategy is in terms of the numbers of link (counting) and node (video/image) sensors and their installation locations. We assume that each time-dependent O-D demand is a random variable, and the variability is measured by the trace of the covariance matrix of the posterior O-D demand estimates. In the proposed model, counting sensors are assumed to be located on links to measure link flows, while video/image sensors are assumed to be located at nodes to measure turning movements. We assume that the cost of a counting sensor is cheaper than that of a video/image sensor. To factor the time duration for which traffic data measurements are available, we add a time duration constraint to the proposed NSLP model that specifies the time duration for which traffic data measurements are available. To study the impact of time duration constraint on the optimal sensor deployment strategy, we compare the optimal sensor deployment strategy using traffic data measurements for the entire time period to the strategy using traffic data measurements for a shorter time period.

A sequential sensor location algorithm that avoids matrix inversions is introduced to solve the proposed NSLP model. Since the costs of link and node sensors are different, the optimal numbers of link and node sensors cannot be simultaneously identified using a simple approach. The proposed algorithm first assumes the number of node sensors in the network as given, and under this scenario selects the sensor deployment strategy under a budget constraint with the lowest variability in the dynamic O-D demand estimates by sequentially adding one sensor (including sensor type and location) at a time to avoid matrix inversions and simplify the computation. The process is repeated for other scenarios with different given number of node sensors in the network. It then compares the selected sensor deployment strategies for various scenarios with given number of node sensors, and selects the optimal sensor deployment strategy as the one with the lowest variability in the dynamic O-D demand estimates.

The remainder of the paper is organized as follows. The next section analyzes the
relationship among the variables considered in the NSLP model. Then, the NSLP model is formulated. Next, a sequential sensor location algorithm is proposed for solving the NSLP model under the multivariate normal distribution assumption, followed by numerical experiments to illustrate the performance of the proposed model and its solution algorithm. The paper concludes with some comments.

**RELATIONSHIPS AMONG VARIABLES CONSIDERED IN THE NSLP MODEL**

The proposed NSLP model incorporates the prior and posterior variability of the dynamic O-D demand. To derive the variability, we first analyze the relationships among the variables considered in the NSLP model, including time-dependent O-D demand, path flows, link flows and node turning movements.

**Relationship between O-D Demand and Path Flow**

Consider the following flow conservation equation:

\[ f_{i,k,t} = p_{i,k,t}d_{i,t} \]  

(1)

where \( d_{i,t} \) is the flow of O-D pair \( i \) with departure time \( t \), \( f_{i,k,t} \) is the number of users between O-D pair \( i \) choosing path \( k \) with departure time \( t \), and \( p_{i,k,t} \) is the proportion of users between O-D pair \( i \) with departure time \( t \) choosing path \( k \).

Define matrices \( D \), \( F_i \), and \( F \) as follows:

\[ D = \begin{bmatrix} d_{1,1}, \ldots, d_{1,t}, \ldots, & d_{i,1}, \ldots, d_{i,t}, \ldots, & \ldots \end{bmatrix}^T \]  

(2)

\[ F_i = \begin{bmatrix} f_{i,1,1}, \ldots, f_{i,1,t}, \ldots, & f_{i,k,1}, \ldots, f_{i,k,t}, \ldots \end{bmatrix}^T \]  

(3)

\[ F = [F_1^T, F_2^T, \ldots, F_i^T, \ldots]^T \]  

(4)

where \( D \) is the vector of all considered time-dependent O-D demands, \( F_i \) is the vector of all time-dependent path flows between O-D pair \( i \) and \( F \) is the vector of all time-dependent path flows.

Define a \( m \times s \) matrix \( P_{i,t} \), where \( m \) is the total number of paths for O-D pair \( i \) with departure time \( t \), and \( s \) is the dimension of \( D \). The \((k,j)^{th}\) element \( P_{i,t}(k,j) \) of \( P_{i,t} \) is defined as:

\[ P_{i,t}(k,j) = \begin{cases} p_{i,k,t} & \text{if } j = (i - 1) \times |T| + t \\ 0 & \text{otherwise} \end{cases} \]  

(5)

where \(|T|\) is the number of time intervals.

Define matrix \( P \) as follows:

\[ P = [P_{1,1}^T, P_{1,2}^T, \ldots, P_{2,1}^T, P_{2,2}^T, \ldots, P_{i,1}^T, P_{i,2}^T, \ldots]^T \]  

(6)

Then, the path flows satisfy the following flow conservation condition:

\[ F = PD \]  

(7)

The time-dependent path flow is the product of the path choice proportion and the time-dependent O-D demand. Note that the path choice proportion is a deterministic variable, which can be obtained by solving the dynamic user equilibrium (DUE) problem. Then, the time-dependent node turning movements and link flows can be derived from the path flows.
Relationship between O-D Demand and Link Flow

Define $v_j$ as the flow on link $j$, and $\varphi_{i,k,t,j}$ as the time-dependent link-path incidence indicator. $\varphi_{i,k,t,j} = 1$, if link $j$ belongs to path $k$ of O-D pair $i$ with departure time $t$, and $\varphi_{i,k,t,j} = 0$, otherwise.

The link flow can be derived from the time-dependent path flows as follow:

$$v_j = \sum_i \sum_k \sum_t \varphi_{i,k,t,j} f_{i,k,t} \quad (8)$$

Define matrices $\varphi_{i,j}$, $\varphi_j$, $\Phi$ and $V$ as follows:

$$\varphi_{i,j} = [\varphi_{i,1,1,j}, \varphi_{i,1,2,j}, \ldots, \varphi_{i,k,1,j}, \varphi_{i,k,2,j}, \ldots] \quad (9)$$

$$\varphi_j = [\varphi_{1,j}, \varphi_{2,j}, \ldots, \varphi_{i,j}, \ldots] \quad (10)$$

$$\Phi = [\varphi_1^T, \varphi_2^T, \ldots, \varphi_j^T, \ldots]^T \quad (11)$$

$$V = [v_1, v_2, \ldots, v_j, \ldots]^T \quad (12)$$

where $V$ is the vector of all considered link flows and $\Phi$ is the corresponding incidence indicator vector. $\varphi_{i,j}$ is the vector of incidence indicators for each time-dependent path of O-D pair $i$ that includes link $j$. $\varphi_j$ is the vector of incidence indicators related to each time-dependent path of each O-D pair that includes link $j$.

By considering an error term $\varepsilon$, the following linear relationship can be expressed between the path flow vector $F$ and the link flow vector $V$:

$$V = \Phi F + \varepsilon = \Phi PD + \varepsilon \quad (13)$$

where $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots)$ are mutually independent random variables with zero mean.

Relationship between O-D Demand and Node Turning Movement

Define $s_{i,a,b,\tau}$ as the number of users traveling from upstream node $a$ to downstream node $b$ connected by node $j$ with departure time $\tau$, and $\psi_{i,k,t,j,a,b}$ is the incidence indicator (i.e., $\psi_{i,k,t,j,a,b} = 1$, if the sub-path made up of nodes $a, j$ and $b$ belongs to path $k$ of O-D pair $i$ with departure time $t$, and $\psi_{i,k,t,j,a,b} = 0$, otherwise). Here, $a \in N_u$ and $b \in N_d$, where $N_u$ is the set of upstream nodes of node $j$ and $N_d$ is the set of downstream nodes of node $j$.

Define a column vector $S_j$ as the set of all the turning movements at node $j$ and a row vector $\Psi_{i,k,t,j}$ as the set of incidence indicators related to path $k$ of O-D pair $i$ with departure time $t$ choosing each turning movement at node $j$. Then, vectors $\psi_{i,j}$, $\psi_j$, $\Psi$, $\Phi$ and $S$ are defined as follow:

$$\Psi_{i,j} = [\psi_{i,j,i,1,1,j}, \psi_{i,j,i,1,2,j}, \ldots, \psi_{i,j,i,k,1,j}, \psi_{i,j,i,k,2,j}, \ldots] \quad (14)$$

$$\psi_j = [\psi_{1,j}, \psi_{2,j}, \ldots, \psi_{i,j}, \ldots] \quad (15)$$

$$\Psi = [\psi_1^T, \psi_2^T, \ldots, \psi_j^T, \ldots]^T \quad (16)$$

$$S = [S_1, S_2, \ldots, S_j, \ldots]^T \quad (17)$$

where $S$ is the vector of all considered turning movements and $\Psi$ is the corresponding incidence indicator vector. $\Psi_{i,j}$ is the vector of incidence indicators related to each
time-dependent path of O-D pair $i$ choosing each turning movement at node $j$. $\Psi_j$ is the vector of incidence indicators related to each time-dependent path of each O-D pair choosing each turning movement at node $j$.

By considering an error term $\eta$, the following linear relationship can be expressed between the path flow vector $F$ and the turning movement vector $S$:

$$S = \psi F + \eta = \psi PD + \eta$$

where $\eta = (\eta_1, \eta_2, \ldots)$ are mutually independent random variables with zero mean.

### Relationships among all of the Considered Variables

According to Eqs. (7), (13) and (18), the random variables in the proposed model can be expressed through the following linear relationships:

$$
\begin{pmatrix}
    D \\
    F \\
    V \\
    S
\end{pmatrix} = 
\begin{pmatrix}
    1 & 0 & 0 \\
    P & 0 & 0 \\
    \phi P & 1 & 0 \\
    \Psi P & 0 & 1
\end{pmatrix}
\begin{pmatrix}
    D \\
    \varepsilon \\
    \eta
\end{pmatrix}
$$

(19)

In this study, since all of the considered variables are treated as random variables, the traffic demands between all time-dependent O-D pairs are assumed to follow multivariate normal (MVN) distributions. This is because these random variables are the outcome of a large number of independent Bernoulli experiments in which the users decide where to travel and which routes to choose. This assumption is similar to assumptions made in past studies (33-34).

Specifically, $D$ are multivariate normal random variables with mean $E(D)$ and variance $\Sigma_D$. $\varepsilon$ are mutually independent normal random variables with mean $E(\varepsilon)$ and variance $\Sigma_\varepsilon$, and $\eta$ are mutually independent normal random variables with mean $E(\eta)$ and variance $\Sigma_\eta$. Based on Eq. (19), the prior variance-covariance matrix $\Sigma_{(D,F,V,S)}$ is:

$$
\Sigma_{(D,F,V,S)} = 
\begin{pmatrix}
    \Sigma_D & \Sigma_D P^T & \Sigma_D (\phi P)^T & \Sigma_D (\Psi P)^T \\
    \Sigma_D P & \Sigma_D P^T & \Sigma_D (\phi P)^T & \Sigma_D (\Psi P)^T \\
    \phi P \Sigma_D & \phi P \Sigma_D P^T & \phi P \Sigma_D (\phi P)^T + \Sigma_\varepsilon & \phi P \Sigma_D (\Psi P)^T \\
    \Psi P \Sigma_D & \Psi P \Sigma_D P^T & \Psi P \Sigma_D (\phi P)^T & \Psi P \Sigma_D (\Psi P)^T + \Sigma_\varepsilon
\end{pmatrix}
$$

(20)

The variance-covariance matrix reflects the randomness of all variables, including time-dependent O-D demands, node turning movements and link flows. Based on the variance-covariance matrix, the correlations among the considered variables can be derived, and these correlations can be used to derive the NSLP model as shown in the next section.

### FORMULATION OF THE NSLP MODELS

The prior distribution $\phi_D$ (i.e., $E(D)$ and $\Sigma_D$ in Eq. (20)) of time-dependent O-D demand is assumed to be obtained using historical O-D data. From Eq. (20), we can derive the prior distribution of all variables. In turn, let $Z$ be a sensor deployment strategy whose counted flows are known to be $Z = z$ (including observed link flows and node turning movements). Then, $\phi_{D|Z=z}$, obtained by updating the counted flows, is the posterior distribution of the time-dependent O-D demand.

The trace $Tr(\Sigma_X)$ of the covariance matrix $\Sigma_X$ has been previously adopted as a measure of variability related to the random vector $X$. Simonelli et al. (13) and Zhou and List...
(24) used the trace of covariance matrix of O-D demand to measure the variability of static O-D demand estimation. We adopt a similar method by using the trace $\text{Tr}(\Sigma_{D|Z=z})$ to represent the variability of the dynamic O-D demand conditional on the counted flows $Z = z$. Since $\text{Tr}(\Sigma_{D|Z=z})$ generally depends on the counted flows $Z = z$, the variability of the posterior random vector $D|Z = z$ can be defined as the average of $\text{Tr}(\Sigma_{D|Z=z})$, as follows:

\[
E[\text{Tr}(\Sigma_{D|Z=z})] = \int_{\Omega_Z} \text{Tr}(\Sigma_{D|Z=z}) \phi_Z \, dz
\]

where $\phi_Z$ is the density function of the random variable $Z$ and $\Omega_Z$ is its domain.

The NSLP can be formulated as the problem of finding the optimal sensor deployment strategy $Z^*$, which minimizes the variability of the posterior random vector $D|Z = z$ in the domain $\Omega_Z$, subject to a budget constraint. Then, the NSLP model for dynamic O-D demand estimation is proposed as follow:

\[
\min \int_{\Omega_Z} \text{Tr}(\Sigma_{D|Z=z}) \phi_Z \, dz \tag{22}
\]

\[
\text{s.t. } b_pL + b_nS \leq b_{\text{max}} \tag{23}
\]

where $b_p$ is the cost of a link sensor, $b_n$ is the cost of a node sensor, $b_{\text{max}}$ is the overall available budget, $L$ is the cardinality of the identified link set, and $S$ is the cardinality of the identified node set.

By optimizing (22) and (23), the resultant optimal heterogeneous sensor deployment strategy depends on the variability of both prior and posterior O-D demand estimates. The posterior estimates are related to the relationships among the considered variables, which depend on the network topology and the users’ travel behaviors. In summary, the optimal heterogeneous sensor deployment strategy obtained by solving the proposed NSLP model incorporates the variability of prior O-D demand estimates, the network topology, and the users’ travel behaviors.

As we have obtained the prior means and variance-covariance matrix of all the variables as shown in Eq. (20), the covariance matrix of the variables can be updated based on some observed variables using the following equations (35, 36) under the assumption of normal distribution:

\[
\Sigma_{XY|Z=z} = \Sigma_{XY} - \Sigma_{XZ}\Sigma^{-1}_{ZZ}\Sigma_{YZ}
\]

where $X$ and $Y$ both refer to the components of $(D, F, V, S)$; $\Sigma_{ZZ}$ is the covariance matrix of the observation $Z$; $\Sigma_{XZ}$ is the covariance matrix of $X$ and $Z$; $\Sigma_{YZ}$ is the covariance matrix of $Y$ and $Z$; $\Sigma_{XY}$ is the covariance matrix of $Y$ and $X$; and $\Sigma_{XY|Z=z}$ is the posterior covariance matrix of $Y$ and $X$.

Under the assumption of multivariate normal distribution, it can be shown that the conditional variance $\Sigma_{XY|Z=z}$ does not depend on the specific counted values of $Z$ (i.e. $z$). Therefore, the trace $\text{Tr}(\Sigma_{D|Z=z})$ does not depend on the actual values of the traffic data measurements, but just on the random sensor deployment strategy $Z$.

Hence, the optimization problem (22) and (23) can be rewritten as:

\[
\min \int_{\Omega_Z} \text{Tr}(\Sigma_{D|Z=z}) \phi_Z \, dz = \text{Tr}(\Sigma_D - \Sigma_{DZ}\Sigma^{-1}_{ZZ}\Sigma_{DZ}) \tag{25}
\]

\[
\text{s.t. } b_pL + b_nS \leq b_{\text{max}} \tag{26}
\]

where $\Sigma_{DZ}$ and $\Sigma_{ZZ}$ are derived from Eq. (20).

In real-world applications, the time duration for which traffic data measurements are available is usually limited to a shorter time period compared to the entire time period of
interest related to determining the dynamic O-D demand. To factor this, the proposed model includes a time duration constraint:

$$\min \int_{\Omega_2} T r(\Sigma_{D_{|Z=z}}) \phi_z d z = T r(\Sigma_D - \Sigma_D Z Z^{-1} \Sigma_D)$$  \hspace{1cm} (27)

s.t. \begin{align*}
    b_p L + b_s N & \leq b_{\text{max}} \\
    \tau_i & \in [\tau_{\text{min}}, \tau_{\text{max}}], \forall i \in Z
\end{align*}

where \( \tau_i \) is the departure time of user at sensor location \( i \); \( \tau_{\text{min}} \) is the start time of the considered time duration for which traffic data measurements are available and \( \tau_{\text{max}} \) is the end time.

Due to the time duration constraint, incidence indicators \( \varphi_{i,k,t,j} \) and \( \psi_{i,k,t,j,a,b} \) are not necessarily 0 or 1 for deriving \( \Sigma_{D_{Z}} \) and \( \Sigma_{Z Z} \) using Eq. (20). In this case, \( \varphi_{i,k,t,j} \) is equal to the proportion of users of path \( k \) of O-D pair \( i \) with departure time \( t \) choosing link \( j \) with departure time \( \tau_j \) which is in the time period \([\tau_{\text{min}}, \tau_{\text{max}}]\). \( \psi_{i,k,t,j,a,b} \) is equal to the proportion of users of path \( k \) of O-D pair \( i \) with departure time \( t \) traveling from upstream node \( a \) to downstream node \( b \) connected by node \( j \) with departure time \( \tau_j \) included in the time period \([\tau_{\text{min}}, \tau_{\text{max}}]\). Given the prior O-D demand, these proportions can be obtained by solving the DUE problem.

We denote the model without the time duration constraint as NSLP-NT model, and the model with the time duration constraint as NSLP-T model.

**THE SEQUENTIAL SENSOR LOCATION ALGORITHM**

As the costs for a link sensor and a node sensor are different, the optimal numbers of link sensors and node sensors cannot be determined simultaneously in a simple manner. Therefore, we first specify a given number of node (or link) sensors in the network, and select the sensor deployment strategy with the lowest variability in the dynamic O-D demand estimates. Then, we compare the selected sensor location strategy for different given number of node (or link) sensors, and choose the optimal strategy as the one with the lowest variability.

To solve the aforementioned NSLP models, calculating the inverse of \( \Sigma_{Z Z} \) (i.e., \( \Sigma_{Z Z}^{-1} \)) requires a large amount of computational effort, especially for large-scale networks since the dimension of \( \Sigma_{Z Z} \) is usually very large. Interestingly, if we sequentially update one observed variable at a time in Eq. (24), it does not involve matrix inverse calculation, because in such a case, \( \Sigma_{D_{Z}} \) is a column vector and \( \Sigma_{Z Z} \) is a scalar (i.e., \( \Sigma_{Z Z} = \sigma_{Z Z} \)). Hence, the proposed NSLP-NT and NSLP-T models can be solved using a sequential sensor location algorithm, summarized using the following eight steps:

**Step 0:** Initialization: Calculate the choice proportion matrices \( \overline{P}, \overline{Q} \) and \( \overline{\Psi} \) by solving the DUE problem.

**Step 1:** Calculate the prior variance and covariance of all considered variables according to Eq. (20).

**Step 2:** Define the maximum number of node sensors \( n_{\text{max}} \) based on the budget constraint \( b_{\text{max}} \), i.e., \( n_{\text{max}} = \text{int}(b_{\text{max}}/b_s) \). Define \( m \) as the number of node sensors, whose initial value is 0. Define \( \lambda_{\text{min}} \) as the minimum value of the objective function, whose initial value is equal to the trace of the prior covariance matrix of the dynamic O-D demand, that is, \( \lambda_{\text{min}} = \text{tr}(\Sigma_D) \). Define \( \overline{Z} \) as the current optimal heterogeneous sensor deployment strategy and \( \overline{Z} \) as the current optimal heterogeneous sensor deployment strategy;
the initial values of \( \tilde{Z} \) and \( Z \) are both set to be null sets.

**Step 3:** Identify the maximum number of link sensors \( l_{\text{max}} \) \( (l_{\text{max}} = \text{int}[(b_{\text{max}} - b_xm)/b_v]) \). Define \( l \) as the number of the identified link sensor locations, \( n \) as the number of the identified node sensor locations and set the initial values of \( l \) and \( n \) to be 0. Define \( \mu \) as the current value of the objective function, whose initial value is \( \mu = \text{tr}(\Sigma_D) \).

**Step 4:** Identify one sensor location and add it to the sensor deployment strategy \( Z \). Define link or node \( Z^* \) as the next sensor location which minimizes the objective function, that is:

\[
Z^* = \arg \min_{Z} \text{Tr}(\Sigma_D - \Sigma_{DZ}^{-1}\Sigma_{DZ})
\]  

(30)

In this step, if \( n < m \) and \( l < l_{\text{max}} \), the new identified sensor can be a link sensor or a node sensor; if \( n = m \) and \( l < l_{\text{max}} \), the new identified sensor can only be a link sensor; and if \( n < m \) and \( l = l_{\text{max}} \), the new identified sensor can only be a node sensor.

**Step 5:** Update the variance, covariance and the current value \( \mu \) of the objective function. According to Eq. (24), the variance and covariance of traffic flows can be updated based on the identified sensor locations using the following formula:

\[
\Sigma_{XYZ}^* = \Sigma_X - \Sigma_{XZ}^{-1}\Sigma_{YZ}/\sigma_{Z^*Z^*}
\]

(31)

where \( X \) and \( Y \) both refer to the components of \( (D, F, V, S) \), \( Z^* \) is the new identified sensor location by Step 4, and \( \sigma_{Z^*Z^*} \), which is a scalar, is the variance of \( Z^* \).

**Step 6:** If the new identified sensor location is a link sensor location, set \( l = l + 1 \); otherwise set \( n = n + 1 \). If \( l = l_{\text{max}} \) and \( n = m \), continue with Step 7. Otherwise, go to Step 4.

**Step 7:** Specify the optimal heterogeneous sensor deployment strategy \( Z \) under the condition in which the defined number of node sensor locations is \( m \). If \( \mu < \lambda_{\text{min}} \), set \( \lambda_{\text{min}} = \mu \), and \( \tilde{Z} = Z \).

**Step 8:** If \( m < n_{\text{max}} \), set \( m = m + 1 \) and initialize the variance and covariance of all variables (using the values based on Step 1); set \( Z \) as a null set and go to Step 3. Otherwise, stop the algorithmic process and specify the optimal heterogeneous sensor deployment strategy \( Z \).

In this algorithm, given the number of node sensors in Step 2, Steps 3-7 identify and update one sensor location at a time. In Step 4, only one sensor location is identified. In this case, because \( \Sigma_{DZ} \) is a column vector and \( \Sigma_{ZZ} \) is a scalar, if the total number of considered links and turning movements in the network is \( a \) and the number of identified sensor locations is \( c \), the number of calculations needed in this step is linearly related to the number of the remaining links and turning movements, i.e., \( a - c \). Since most of the calculations in this algorithm are involved in Step 4, the computational time of solving the NSLP model is linear with respect to the number of links and turning movements in the network.

Note that more than one sensor location may be identified in some cases in Step 4 of the proposed algorithm as these locations lead to the same covariance matrix reduction of O-D demand. The proposed algorithm does not determine which one of them should be considered for the optimal sensor location deployment strategy. In this case, the algorithm needs to first identify all the sensor deployment strategies for each such sensor location with the same covariance matrix reduction of O-D demand and choose the optimal strategy with the lowest variability. This would entail additional computational effort.
NUMERICAL EXPERIMENTS

In this section, we apply the proposed NSLP models and algorithm to a network from a part of Chennai, India as shown in Figure 1, which includes 11 traffic analysis zones (100 O-D pairs), 70 nodes, and 141 directed links. The O-D matrix is time-dependent, and specified in 15-minute intervals. The entire analysis period is two hours (from 8:00 a.m. to 10:00 a.m.); hence the time-dependent O-D matrix has 800 records (8 time intervals for 100 O-D pairs).

The prior O-D demand is established from the historical data available for this network. The total available budget is assumed to be 300. The costs for a node sensor and link sensor are assumed to be 50 and 15, respectively. The numerical experiments were conducted using DYNASMART-P 1.3.0, and the prior traffic data measurements for all links and nodes are obtained by solving the DUE problem based on the prior O-D demand. The DUE paths in DYNASMART are solved by a time-dependent shortest path (TDSP) method in each iteration, and all generated TDSPs are equilibrated for each departure time using the method of successive averages.

FIGURE 1 Network from a part of Chennai, India.

Figure 2 shows the plots of the traces of the covariance matrices of the dynamic O-D demand estimation for the NSLP-NT model using the proposed sequential algorithm under different given number of node sensors (labeled NS# in the figure). Figure 3 shows the corresponding plots for the NSLP-T model, in which the time duration for which traffic data measurements are available is set to be one hour, from 9:00 a.m. to 10:00 a.m. As illustrated by Figures 2 and 3, node sensor locations are identified first in each case. Before a new sensor location is identified, the prior variance-covariance matrix is updated using Eq. (31) based on the identified sensor locations. As shown in Figures 2 and 3, the traces decrease
with each added sensor. More sensors imply that more observed information can be collected to update the variance-covariance matrix, which can reduce the variability of the dynamic O-D demand estimates. The traces decrease more rapidly when a node sensor is added rather than a link sensor, because a node sensor can detect several turning movements (for example, 12 turning movements at a traditional four-way intersection) and thus provide a greater amount of updated information compared to a link sensor. Hence, as the NS# increases, the traces decrease more rapidly as each node sensor is added.

Table 1 shows the optimal sensor deployment strategies for the NSLP-NT model and Table 2 shows the optimal strategies for the NSLP-T model. In terms of the notation in Tables 1 and 2, for example, 7(n) represents a node sensor located at node 7; 16-17 represents a link sensor on link 16-17 with upstream node 16 and downstream node 17.

As illustrated by Tables 1 and 2, under the budget constraint, though a node sensor can provide a greater amount of updated information compared to a link sensor, more node sensors cannot reduce the variability of the O-D demand estimates. Thereby, though the maximum number of node sensors is 6, the number of node sensors in the optimal heterogeneous sensor deployment strategy for the NSLP-NT model is 3 because it leads to the lowest objective function value (bolded in the table). When there are 6 node sensors in the network, the allowed number of link sensors is 0. However, in the optimal heterogeneous sensor deployment strategy, the number of link sensors is 8, though the number of node sensors is only 3. This is because a link sensor is much cheaper than a node sensor, and more node sensors can imply fewer link sensors, leading to a tradeoff in terms of the desirable numbers of each sensor type. The number of node sensors in the optimal heterogeneous sensor deployment strategy for the NSLP-T model is 5. This higher value compared to the NSLP-NT case is because the more constrained context due to the time duration constraint under NSLP-T favors the use of more node sensors to elicit more information. In summary, the optimal numbers for both the link sensors and node sensors can be determined using the proposed algorithm.

The results in Tables 1 and 2 also show that the locations of node sensors are identified first by the proposed algorithm. This is because a node sensor can provide a greater amount of updated information compared to a link sensor, and is hence more effective in reducing the variability of the O-D demand estimation.

In addition, as illustrated by Tables 1 and 2, the optimal heterogeneous sensor deployment strategy obtained for the NSLP-T model is different from that of the NSLP-NT model. This implies that considering the time duration for which traffic data measurements are available can lead to a different optimal sensor deployment strategy. If traffic data measurements are available for the entire time period of interest related to determining the dynamic O-D demand, each sensor location identified by the NSLP-NT model can provide a greater amount of updated information compared to non-selected sensor locations. However, if the time duration for which traffic data measurements are available is considered, each sensor location identified by the NSLP-NT model may not provide a greater amount of updated information compared to the non-selected ones due to the time-varying characteristics of the amount of updated information that can be provided by each sensor location. Since O-D demand and users’ travel behavior are time-dependent, the amount of updated information that can be provided by each sensor location is different for each time.
period. Hence, the optimal heterogeneous sensor deployment strategy obtained by the NSLP-T model is different from that of the NSLP-NT model.

**FIGURE 2** Traces of covariance matrices of dynamic O-D demand estimation after updating each identified sensor location with different given number of node sensors, without considering the time duration constraint.

**FIGURE 3** Traces of covariance matrices of dynamic O-D demand estimation after updating each identified sensor location with different given number of node sensors, considering the time duration constraint.

The objective function value of the NSLP-NT model is lower than that of the NSLP-T model as it is a less constrained problem. This implies that the posterior variance of variables in the NSLP-T model is larger than that in the NSLP-NT model. This is because in the NSLP-NT model the updated information is collected for the entire time period of interest related to determining the dynamic O-D demands, while the updated information in the
NSLP-T model is collected for only the time duration specified. Hence, the NSLP-T model has less observed information to update the posterior variance which leads to a larger variability in the dynamic O-D demand estimates as more observed information implies lower variability of the O-D demand estimates (as illustrated in Figures 2 and 3).

**TABLE 1 Sensor Deployment Strategies for the NSLP-NT Model**

<table>
<thead>
<tr>
<th>Number of node sensors</th>
<th>Sensor deployment strategy</th>
<th>Objective function value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>16-17, 22-18, 13-11, 31-3, 3-31, 11-13, 64-7, 29-2, 44-1, 47-19, 10-13, 67-2, 19-47, 7-64, 36-7, 7-36, 1-44, 60-14, 14-60, 10-11</td>
<td>25351.54</td>
</tr>
<tr>
<td>1</td>
<td>7(n), 16-17, 10-7, 7-36, 13-11, 22-18, 31-3, 10-11, 3-31, 11-13, 44-1, 47-19, 29-2, 10-13, 67-2, 19-47, 14-60</td>
<td>24724.28</td>
</tr>
<tr>
<td>2</td>
<td>7(n), 3(n), 1-2, 16-14, 19-1, 22-18, 18-16, 3-31, 11-13, 14-16, 31-3, 7-36, 18-19, 14-13, 10-11</td>
<td>21507.6</td>
</tr>
<tr>
<td>3</td>
<td>7(n), 3(n), 1(n), 15, 12, 3-31, 5-3, 7-36, 67-2, 19-18, 1-44, 19-1, 16-17</td>
<td><strong>21338.33</strong></td>
</tr>
<tr>
<td>4</td>
<td>7(n), 3(n), 1(n), 16(n), 18-16, 3-31, 5-3, 7-36, 22-18, 19-1</td>
<td>21682.16</td>
</tr>
<tr>
<td>5</td>
<td>7(n), 3(n), 1(n), 16(n), 2(n), 5-3, 3-31, 7-36</td>
<td>21591.57</td>
</tr>
<tr>
<td>6</td>
<td>7(n), 3(n), 1(n), 16(n), 2(n), 10(n)</td>
<td>21777.34</td>
</tr>
</tbody>
</table>

**TABLE 2 Sensor Deployment Strategies for the NSLP-T Model**

<table>
<thead>
<tr>
<th>Number of node sensors</th>
<th>Sensor deployment strategy</th>
<th>Objective function value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>16-17, 13-11, 31-3, 11-13, 22-18, 3-31, 1-44, 64-7, 44-1, 10-11, 19-47, 10-13, 29-2, 36-7, 7-36, 67-2, 60-14, 14-60, 47-19, 2-29</td>
<td>25352.05</td>
</tr>
<tr>
<td>1</td>
<td>7(n), 16-17, 64-7, 7-36, 13-11, 19-47, 1-44, 11-13, 36-7, 3-31, 60-14, 7-10, 5-7, 31-3, 44-1, 10-13, 22-18</td>
<td>25074.89</td>
</tr>
<tr>
<td>2</td>
<td>7(n), 2(n), 16-17, 2-67, 2-29, 64-7, 7-36, 13-11, 1-2, 67-2, 29-2, 11-13, 19-47, 36-7, 60-14</td>
<td>24851.41</td>
</tr>
<tr>
<td>3</td>
<td>7(n), 2(n), 14(n), 16-17, 2-67, 2-29, 64-7, 14-60, 11-13, 1-2, 60-14, 67-2, 22-18</td>
<td>24761.18</td>
</tr>
<tr>
<td>4</td>
<td>7(n), 2(n), 14(n), 16(n), 14-16, 2-67, 2-29, 64-7, 14-60, 16-17</td>
<td>24630.79</td>
</tr>
<tr>
<td>5</td>
<td>7(n), 2(n), 14(n), 16(n), 1(n), 44-1, 60-14, 1-2</td>
<td><strong>24308.59</strong></td>
</tr>
<tr>
<td>6</td>
<td>7(n), 2(n), 14(n), 16(n), 1(n), 10(n)</td>
<td>24603.31</td>
</tr>
</tbody>
</table>

To further compare the NSLP-T and the NSLP-NT models, we use the optimal sensor deployment strategy from the NSLP-NT model to collect traffic data measurements, but now with the time duration constraint used for the NSLP-T model. Figure 4 shows how the traces of the covariance matrices of the dynamic O-D demand estimates of the two models change after updating the identified sensor location one-at-a-time using the proposed sequential algorithm. Note that compared to the NSLP-NT model, the trace of the covariance matrix of the dynamic O-D demand estimation under the NSLP-T model decreases more rapidly and
the optimal value is lower (the traces in NSLP-T and NSLP-NT models are 24308.59 and 25123.47, respectively). This demonstrates the superior accuracy and performance of the NSLP-T model. That is, the optimal sensor deployment strategy by the NSLP-T model performs better than the optimal sensor deployment strategy by the NSLP-NT model with the consideration of the same time duration constraint used in the NSLP-T model.

![Figure 4](image)

**FIGURE 4** Traces of covariance matrices of the dynamic O-D demand estimates for the two models after updating the sensor locations under the same time duration constraint.

As shown in Tables 1 and 2, and Figures 2 to 4, the proposed algorithm can identify sensor locations with higher priorities based on the amount of updated information they can provide. Node sensor locations are given higher priorities compared to link sensor locations because of their ability to provide a greater amount of information. In addition, the proposed algorithm can identify the optimal sensor location strategy in terms of sensor types, and their numbers and locations for dynamic O-D demand estimation under a budget constraint. The results also show that the optimal sensor location strategy can change significantly with the consideration of the time duration for which traffic data measurements are available. This illustrates that the proposed NSLP-T model can capture the time-varying characteristics of the amount of information provided by each sensor location compared to the NSLP-NT model.

**CONCLUDING COMMENTS**

This study proposes a network sensor location problem model to determine the optimal heterogeneous sensor deployment strategy for the dynamic O-D demand estimation problem. By maximizing the quality or minimizing the variability of the O-D demand estimates under a given budget constraint, the proposed model can be used to determine the optimal link (counting) and node (video/image) sensors numbers and their installation locations. In the proposed model, the trace of the covariance matrix of the posterior O-D demand estimates is adopted as a measure of the variability of the O-D demand estimates. Under the multivariate normal distribution assumption for the prior O-D demand, the trace of the covariance matrix of the posterior O-D demand estimates does not depend on the actual values of traffic data.
measurements, but only on the sensor locations. The time duration for which traffic data measurements are available is constrained in the proposed model to factor that traffic data are usually collected for a shorter time period in practice rather than the entire time period of interest (such as the peak period) in the context of determining the optimal sensor deployment strategy. A sequential sensor location algorithm that avoids computing the inverse matrix is proposed to simplify the computation and facilitate the application of the proposed model in real-world networks.

Numerical experiments are used to demonstrate the performance of the proposed NSLP model and its solution algorithm. Results show that the trace of the covariance matrix of the dynamic O-D demand estimates decreases when each sensor is added using the proposed algorithm. The traces of the covariance matrices decrease even more rapidly when adding a node sensor compared to a link sensor, because a node sensor collects more information than a link sensor. However, under the budget constraint, more node sensors cannot further reduce the variability of the O-D demand estimates. This is because the much cheaper cost of a link sensor compared to a node sensor entails fewer link sensors when more node sensors are selected. Hence, the optimal sensor deployment strategy may not have the most number of node sensors.

The trace of the covariance matrix of the dynamic O-D demand estimates in the NSLP-T model is larger than that in the NSLP-NT model. However, when the updated information collected using the sensor deployment strategy of each model is restricted by the time duration constraint, the trace of the covariance matrix of the dynamic O-D demand estimates in the NSLP-T model deceases more rapidly and is smaller than that in the NSLP-NT model.

The proposed model and algorithm can be used to develop effective strategies for identifying sensors with fixed locations in terms of their types, numbers and locations in the context of dynamic O-D demand estimation under a budget constraint. A potential future research direction is to leverage mobile sensors (for example, GPS, AVI or global system for mobile communications (GSM)) in the NSLP model to enhance the quality of the dynamic O-D demand estimates.

ACKNOWLEDGMENTS

This study is based on research supported by the NEXTRANS Center, the USDOT Region 5 University Transportation Center at Purdue University. It is also supported by the National Natural Science Foundation of China (No. 51178110, No. 51378119 and No. 61304205) and by the Scientific Research Foundation of Graduate School of Southeast University (No. 3221005743). The authors are solely responsible for the contents of this paper.

REFERENCES


