USDOT Region V Regional University Transportation Center Final Report

Determination of Network Origin-Destination Matrices Using Partial Link Traffic Counts and Virtual Sensor Information in an Integrated Corridor Management Framework

By

Shou-Ren Hu
Associate Professor
Department of Transportation and Communication Management Science
National Cheng Kung University, Taiwan
shouren@mail.ncku.edu.tw

and

Han-Tsung Liou
Graduate Research Assistant
Department of Transportation and Communication Management Science
National Cheng Kung University, Taiwan
iroya.liou@gmail.com

and

Srinivas Peeta
Professor
School of Civil Engineering, Purdue University
peeta@purdue.edu

and

Hong Zheng
Postdoctoral Research Associate
NEXTRANS Center, Purdue University
zheng225@purdue.edu

DISCLAIMER

USDOT Region V Regional University Transportation Center Final Report

TECHNICAL SUMMARY

Title

Determination of Network Origin-Destination Matrices Using Partial Link Traffic Counts and Virtual Sensor Information in an Integrated Corridor Management Framework

Introduction

Trip origin-destination (O-D) demand matrices are critical components in transportation network modeling, and provide essential information on trip distributions and corresponding spatiotemporal traffic patterns in traffic zones in vehicular networks. Trip O-D matrices also reflect traffic loadings and flow intensities in transportation networks, and are crucial inputs in determining short-term traffic control schemes and long-term transportation improvement programs, as well as offline transportation planning and online traffic management. Trip O-D demand matrices have traditionally been estimated by conducting household surveys or roadside interviews; however, this is infeasible because of the high cost and data recording errors involved. Inferring network O-D demand matrices using corresponding link flows is an effective alternative approach, because link flows, which are a set of traffic flows associated with the vehicular trip distributions of different O-D pairs, are easily obtained. Past studies on the estimation of network O-D demand matrices using link flow information have generally assumed that link flows are fully observable. However, in practice, highway agencies face budget constraints in implementing comprehensive sensor deployment plans, and assuming the full observability of link flows is unreasonable. Because of the rapid development of information and communication technologies (ICTs), applications of advanced sensor technologies to traffic management and operation have become widespread and essential. As a result, determining the strategic deployment of traffic sensors to obtain necessary traffic information for network O-D demand estimation has become crucial in transportation network research.

The performance of a network O-D demand estimation model is strongly dependent on the quantity and quality of traffic data collected by different types of traffic sensors. The purpose of the Network Sensor Location Problem is to determine the optimal, minimum number of required traffic sensors and identify their corresponding installation locations, especially under the limited budget constraints of highway agencies. The collected partial link and path flow data are crucial inputs used to estimate corresponding O-D demands in a vehicular network. The strategic deployment of heterogeneous traffic sensors for network O-D demand estimation is a critical subject in transportation network science. The purpose of this study is to develop an integrated heterogeneous sensor deployment model to estimate network O-D demands. One of the unique aspects of the proposed model framework is that it does not require the unreasonable assumption of known prior O-D demand information, turning proportions, or route choice

probabilities, enabling the network O-D demand and path flow estimation problems to be more practically traceable.

Findings

This study addresses the two primary objectives:

1. Propose an effective generalized sensor location model for sensor location flow-observability and sensor location flow-estimation problems.

2. Give an assumption-free, link-based network O-D demands estimation formulation by leveraging flow information provided by different sensor sources.

This research proposes a "double dummy variable" concept to solve the heterogeneous sensor deployment problem for a vehicular network captured by its link-node incidence matrix. A generalized sensor location model was developed to simultaneously determine the optimal number and installation locations for both passive- and active-type sensors. The optimal sensor location policy is further applied to solve the network O-D demands estimation problem using a link-based flow estimation approach. The proposed integrated sensor location model takes full advantage of strategic link flow information provided by traditional vehicle detectors, and partial path flow information given by virtual sensors. The theoretical background and mathematical properties of the proposed model framework are elaborated. The major contribution of this research is the illustration of an integrated model framework for optimal heterogeneous sensor deployment policy, and its potential to the estimation of network O-D demands. One of the unique aspects of the proposed model framework is that it does not require the unreasonable assumption of known prior O–D demand information, turning proportions, or route choice probabilities, enabling the network O–D demand and path flow estimation problems to be more practically traceable.

Recommendations

This study solved the sensor location and network O-D demand estimation problems in two steps. The first step focuses on the generalized sensor location model, and the second step focuses on the O-D demand estimation model. These two steps are independent of each other. Two future research directions are proposed. First, bi-level programming, in which the upper level is a heterogeneous traffic sensor location model and the lower level is an O-D demand estimation model, can be studied based on the correlation between the upper and lower levels. Second, the heterogeneous traffic sensor location model can be integrated with the network O-D demand estimation model in a single-step framework. A bi-level, integrated model framework would be more straightforward and useful in practical applications.

Contacts
For more information:

Dr. Srinivas Peeta **NEXTRANS Center**
Principal Investigator Purdue University - Discovery Park
Professor of Civil Engineering, & Director 3000 Kent Ave

NEXTRANS Center, Purdue University West Lafayette, IN 47906
Ph: (765) 496 9726
Fax: (765) 807 3123 nextrans@purdue.edu
peeta@purdue.edu (765) 496-9729
www.cobweb.ecn.purdue.edu/~peeta/ (765) 807-3123 Fax
www.purdue.edu/dp/nextrans

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

1.1   *Background and motivation*

Trip origin–destination (O–D) demand matrices are critical components in transportation network modeling, and provide essential information on trip distributions and corresponding spatiotemporal traffic patterns in traffic zones in vehicular networks. Trip O–D matrices also reflect traffic loadings and flow intensities in transportation networks, and are crucial inputs in determining short-term traffic control schemes and long-term transportation improvement programs, as well as offline transportation planning and online traffic management. Trip O–D demand matrices have traditionally been estimated by conducting household surveys or roadside interviews; however, this is unfeasible because of the high cost and data recording errors involved. Inferring network O–D demand matrices using corresponding link flows is an effective alternative approach, because link flows, which are a set of traffic flows associated with the vehicular trip distributions of different O–D pairs, are easily obtained. Past studies on the estimation of network O–D demand matrices using link flow information have generally assumed that link flows are fully observable. However, in practice, highway agencies face budget constraints in implementing comprehensive sensor deployment plans, and assuming the full observability of link flows is unreasonable. Because of the rapid development of information and communication technologies (ICTs), applications of advanced sensor technologies to traffic management and operation have become widespread and essential. As a result, determining the strategic deployment of traffic sensors to obtain necessary traffic information for network O–D demand estimation has become crucial in transportation network research.

1.2   Study objectives

The performance of a network O–D demand estimation model is strongly dependent on the quantity and quality of traffic data collected by different types of traffic sensors. The purpose of the network sensor location problem (NSLP) is to determine the optimal, minimum number of required traffic sensors and identify their corresponding installation locations, especially under the limited budget constraints of highway agencies. The collected partial link and path flow data are crucial inputs used to estimate corresponding O–D demands in a vehicular network. The strategic deployment of heterogeneous traffic sensors for network O–D demand estimation is a critical subject in transportation network science. The purpose of this study was to develop an integrated heterogeneous sensor deployment model to estimate network O–D demands. One of the unique aspects of the proposed model framework is that it does not require the unreasonable assumption of known prior O–D demand information, turning proportions, or route choice probabilities, enabling the network O–D demand and path flow estimation problems to be more practically traceable.

1.3   Organization of the research

The remainder of the research is organized as follows. Chapter 2 discusses the background of the problem and proposes the generalized sensor location model. Chapter 3 presents the O–D demand estimation model, which is based on link and path flow information provided by vehicle detectors (VDs) and license plate recognition (LPR), respectively. Chapter 4 provides the results of the numerical analysis when using the proposed generalized sensor location model and O–D demand estimation model in a hypothetical network and a simplified real road network. Finally, in Chapter 5, we summarize this paper by providing insight and future research directions.

CHAPTER 2. AN GENERALIZED SENSOR LOCATION MODEL

This chapter introduces an generalized sensor location model for the estimation of network O-D demands estimation. Section 2.1 states the background of the sensor location problem. Section 2.2 addresses the passive-type sensor location model. Section 2.3 discusses the active-type sensor location model. Section 2.4 introduces the integrated heterogeneous sensor location by accommodating the various traffic flow information of different sensor sources.

2.1   *Problem statement*

To develop a generalized network sensor location model, this study considered both passive-type and active-type sensor location models. The two location models were integrated using a specially designed double 0-1 dummy variable concept to infer and/or estimate network flow.

Before illustrating the generalized network sensor location model, the problem of multiple solutions is discussed. Table 1 shows the feasible solutions of an NSLP required to enable the full observability of link flows (Hu et al., 2009). The demonstration network in Table 1 consists of 9 nodes and 16 unidirectional links, where Nodes 1, 3, 7, and 9 are centroid nodes and Nodes 2, 4, 5, 6, 8 are noncentroid or intermediate nodes (Ng, 2012).

To fully observe link flows in this hypothetical network, the minimal number of links to be equipped with traffic sensors, as shown in Table 1 those links with solid arrows, is 11. However, as depicted by the four feasible solutions in Scenarios A through D, the 11 sensors could be installed in multiple configurations. To connect to Node 5, Scenario A requires three deployed sensors, Scenario B requires four deployed sensors, Scenario C requires five deployed sensors, and Scenario D requires six deployed sensors.

These numbers differ because the number of sensors is strongly dependent on the sensor deployment conditions of Node 5's upstream and downstream nodes. The dependent relationship is a *chain*. Similar to certain sensor location models, including the link-path incidence and the O-D/path/link incidence matrices, this matrix was constructed based on the spatial relationship between upstream and downstream nodes for each path or O-D pair, enabling the dependent chain to be systematically traced using a pre-specified incidence matrix. However, as previously discussed, such an incidence matrix is difficult to obtain in practice because of the path enumeration problem. On the other hand, sensor location models based on link-node incidence matrices cannot directly capture sensor deployment conditions on specific links for an intermediate node's upstream and downstream nodes; a specific technique is required to manage this situation. This study adopted the link-node incidence matrix approach to easily obtain sensor deployment conditions based on the network's topology. This enabled the degree conditions of each node to be individually investigated and used in constructing respective flow conservation constraints based on a specially designed double 0-1 dummy variable to solve the chain problem.

Table 1. Available solutions for full link flow observability.
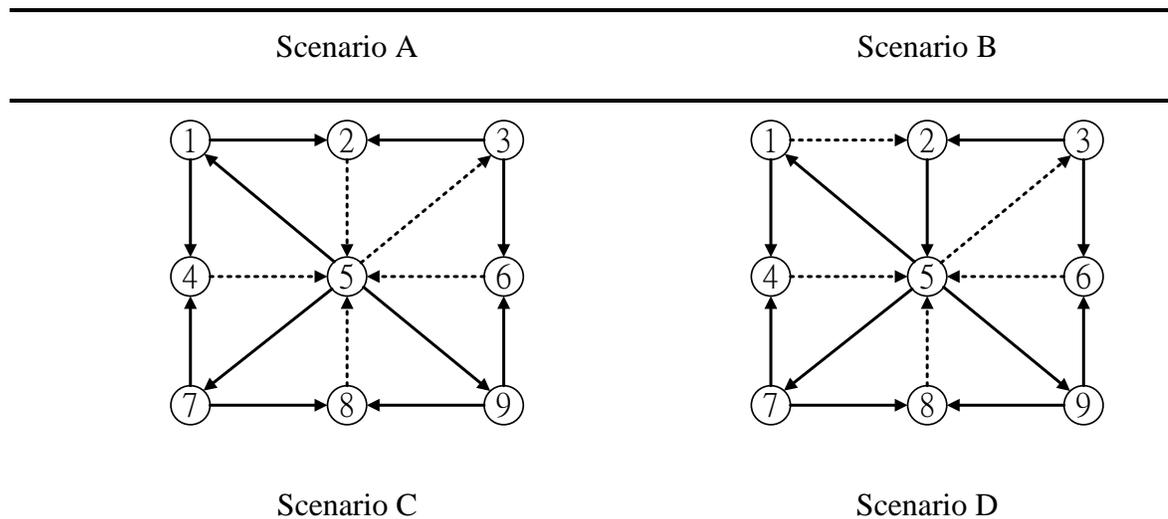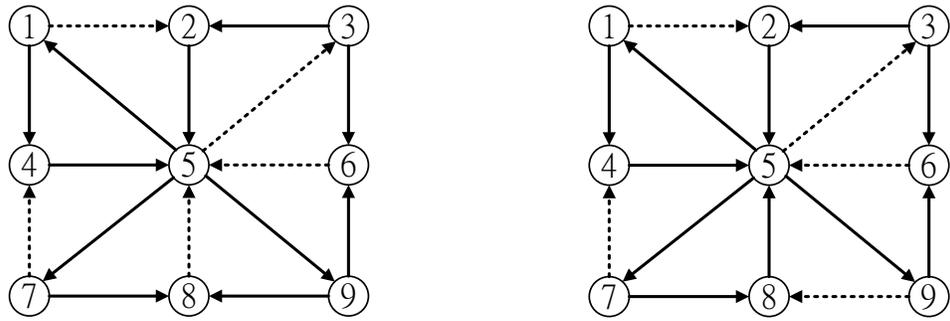
Scenario A	Scenario B

Scenario C	Scenario D

Note: the solid arrow denotes a sensor equipped link; the dotted arrow denotes an unequipped link.

2.2 *Passive-type sensor location model*

The passive-type sensors considered in the passive-type sensor location model were VDs, which can be used to collect link traffic flows or counts. The ideal passive-type sensor location model was based on the degree constraint and link flow conservation rule of the nodes. To determine the link flow conservation rule of an intermediate node (Ng, 2012), the total input flow must be equal to total output flow at that node. In the network in Fig. 1, for example, the summation of link flows in Links 1-3 and 2-3 is equal to that of Links 3-4 and 3-5, and one link flow can be inferred by the flow information contained in the other three links. The flow conservation condition for Node 3 is shown in Eq. 1.
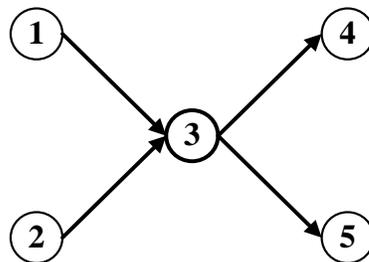
Fig. 1. Example network 1.

$$x_{13} + x_{23} = x_{34} + x_{35} \, , \tag{1}$$

where x_{ij} is the link flow from node i to node j.

Eq. 1 can be augmented in Eq. 2 by incorporating a 0-1 variable, and Eq. 2 is calculated according to the flow conservation rule.

$$x_{13} \cdot 1 + x_{23} \cdot 1 = x_{34} \cdot 1 + x_{35} \cdot 1 \tag{2}$$

If the flow conservation rule causes each element of Eq. 2 to move arbitrarily from the left-hand side (LHS) to the right-hand side (RHS), the flow conservation rule remains valid. Based on Eq. 2, it is obvious that an arbitrary element x can be calculated according to the remaining elements. If x_{35} is selected as a particular element, then Eq. 2 can be converted into Eq. 3. Based on Eq. 3, it is also clear that if x_{13}, x_{23}, x_{34} are observed by traffic sensors, then x_{35} can be calculated according to the flow conservation rule.

$$x_{13} \cdot 1 + x_{23} \cdot 1 - x_{34} \cdot 1 = x_{35} \cdot 1 \tag{3}$$

In other words, to guarantee full link flow observability for the simple network shown in Fig. 1, the number of deployed sensors must be three (one link is an unequipped link). If Eq. 3 does not include flow information, then the flow information can be formulated using 0-1 binary variables: three red 1s and one blue 1 (see Eq. 4).

$$[1,1,1] = 1 \tag{4}$$

Eq. 4 represents the network topology condition for the full observability of a link flow at a node; flow information and signs (plus or minus) can be neglected in such a formulation. Eq. 4 also indicates that the size of the LHS is dependent on the connectivity of a node, and the size of the RHS is always 1. The concepts of in-degree and out-degree

were introduced and incorporated into Eq. 4, to satisfy the new constraint, referred to as the "*degree constraint*" (see Eq. 5).

$$1+1+1 = \mid ID_3 \mid + \mid OD_3 \mid -1 = 2+2-1 \tag{5}$$

where,

$\mid ID_3 \mid$: the size of in-degree for Node 3;

$\mid OD_3 \mid$: the size of out-degree for Node 3;

For Node 3, the in-degree is 2 and the out-degree is 2, and the total degree value is 4. Eq. 5 yields the relationship between equipped links (three red 1s), the unequipped link (one blue 1), and the sizes of the in-degree and out-degree. Based on the flow conservation rule and the degree constraint, we can conclude that the number of deployed sensors for an intermediate node is equal to the total degree value minus 1. This is the definition of the degree constraint (see Eq. 6).

$$DC_j : \mid ID_j \mid + \mid OD_j \mid -1 \tag{6}$$

where,

DC_j : degree constraint for Node *j*;

$\mid ID_j \mid$: the size of in-degree for Node *j*;

$\mid OD_j \mid$: the size of out-degree for Node *j*.

By using the degree constraint and the flow conservation rule to obtain the minimal number of deployed sensors in the passive-type sensor location model, the "*reduplication problem*" may arise. This is when two neighboring nodes share the same link as an equipped link, resulting in overestimation of the number of required sensors for the full observability of link flows. For example, if Link 3-6 in the simple network in Fig. 2 is chosen as an unequipped link, then the degree constraint is satisfied for both Node 3 and Node 6. However, if Link 3-6 is assigned as an unequipped link for Node 3, then a

different link must be selected as the unequipped link for Node 6; thus, a systematic mechanism is required to avert the selection of Link 3-6 as an unequipped link for Node 6. Relationships between neighboring nodes related to equipped or unequipped links are "*series chains*." Using a single 0-1 binary variable to determine sensor deployment conditions for a specific link is inadequate to describe these chains, and the reduplication problem may arise.
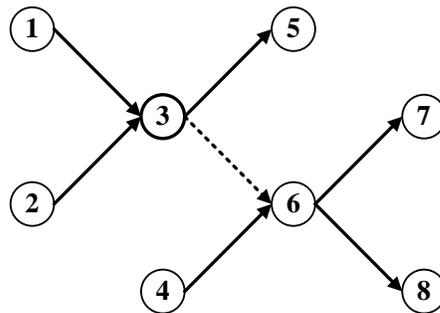
Fig. 2. Example network 2 for the reduplication problem.

Note: the solid arrow denotes a sensor equipped link; the dotted arrow denotes an unequipped link.

To solve the reduplication problem, *double 0-1 binary variables* were introduced to break chains from a set of nodes. Specifically, each link was labeled with two 0-1 binary variables instead of a single 0-1 binary variable. The definitions of the double 0-1 binary variables are described in Eqs. 7 and 8.

$$b(i)_{ij} = \begin{cases} 1, \text{for Node } i, \text{ the flow on Link } i\text{-}j \text{ can be collected by sensor or inferable} \\ 0, \text{for Node } i, \text{ Link } i\text{-}j \text{ is an unequipped link} \end{cases} ; \qquad (7)$$

$$b(j)_{ij} = \begin{cases} 1, \text{for Node } j, \text{ the flow on Link } i\text{-}j \text{ can be collected by sensor or inferable} \\ 0, \text{for Node } j, \text{ Link } i\text{-}j \text{ is an unequipped link} \end{cases} \quad . \quad (8)$$

In Eqs. 7 and 8, a specific Link i-j in a target network is characterized by double 0-1 binary variables, $b(i)_{ij}$ and $b(j)_{ij}$; the first variable denotes the tail node i and the second variable denotes the head node j. Thereby, from a sensor deployment perspective four possible combinations of the double dummy binary variables for a specific Link i-j can be identified:

1)   When $b(i)_{i,j} = 1$ and $b(j)_{i,j} = 1$ holds, for Node i, Link i-j is an equipped link; for Node j, Link i-j is also an equipped link. Since Link i-j is equipped with a sensor, for the head Node j, it can choose one of its remaining adjacent links as an unequipped link.

2)   When $b(i)_{i,j} = 0$ and $b(j)_{i,j} = 1$ holds, it means that for Node i, Link i-j is an unequipped link; for Node j, flow on Link i-j is inferable. Since flow on Link i-j can be inferred, for the head Node j it can choose one of its remaining adjacent links as an unequipped link except for Link i-j.

3)   When $b(i)_{i,j} = 1$ and $b(j)_{i,j} = 0$ holds, this scenario is similar to that of $b(i)_{i,j} = 0$ and $b(j)_{i,j} = 1$. For the tail Node i, flow on Link i-j is inferable and it can choose one of its remaining adjacent links as an unequipped link. For the head Node j, Link i-j is an unequipped link but flow on Link i-j can be inferred.

4)   If $b(i)_{ij} = b(j)_{ij} = 0$, this scenario is prohibited due to the reduplication problem. For Node i and Node j, they cannot simultaneously select Link i-j as an unequipped link.

In summary, three possible results can be obtained: 1) if $b(i)_{i,j} = 1$ and $b(j)_{i,j} = 1$, then Link i-j is an equipped link; 2) if $b(i)_{i,j} = 1$, $b(j)_{i,j} = 0$ or $b(i)_{i,j} = 0$, $b(j)_{i,j} = 1$,

then Link *i-j* is an unequipped but flow-inferable link; and 3) if $b(i)_{i,j} = 0$ and $b(j)_{i,j} = 0$, this is a prohibited scenario which will be avoided in later modeling process.

When Eqs. 7 and 8 are equal to one; it means that they will give positive one into the degree constraint (see Eq. 6). Based on the designated double 0-1 binary variables, degree constraints can be independently set up for each intermediate node and the reduplication problem can be avoided. Table 2 shows a set of feasible solutions for Example network 2 (Fig. 2), where Link 3-6 is an unequipped link for Node 3, and Link 6-8 is an unequipped link for Node 6; this set of feasible solutions is not prone to the reduplication problem. Accordingly, we developed a mathematical program to minimize the number of passive-type sensors required for full observability of link flows, and the constraints are essentially node-degree constraints captured by a set of double 0-1 binary variables.

Table 2. A Set of feasible solutions for example network 2.

Node #	Value of double 0-1 binary variable	Degree constraint
Node 3	$b(3)_{1,3} = 1, b(3)_{2,3} = 1,$ $b(3)_{3,5} = 1, b(3)_{3,6} = 0$	$DC_3 = b(3)_{1,3} + b(3)_{2,3} + b(3)_{3,5} + b(3)_{3,6}$ $= \mid ID_3 \mid + \mid OD_3 \mid -1 = 3$
Node 6	$b(6)_{3,6} = 1, b(6)_{4,6} = 1,$ $b(6)_{6,7} = 1, b(6)_{6,8} = 0$	$DC_6 = b(6)_{3,6} + b(6)_{4,6} + b(6)_{6,7} + b(6)_{6,8}$ $= \mid ID_6 \mid + \mid OD_6 \mid -1 = 3$

The passive-type sensor location model is formulated as follows:

$$\text{Min} \sum_{(i,j) \in A} ps_{i,j}$$

(9)

s.t.

$$\sum_{i}^{|I|} b(j)_{i,j} + \sum_{k}^{|K|} b(j)_{j,k} = |ID_j| + |OD_j| - 1, \forall j \in N - R - S \tag{10}$$

$$b(i)_{i,j} + b(j)_{i,j} \geq 1, \forall i, j \in N - R - S, (i,j) \in A \tag{11}$$

$$b(i)_{j,i} + b(j)_{i,j} \geq 1, \forall i, j \in N - R - S, (i,j) \in A \tag{12}$$

$$b(i)_{i,j} + b(j)_{j,i} \geq 1, \forall i, j \in N - R - S, (i,j) \in A \tag{13}$$

$$b(i)_{i,j} + b(j)_{i,j} - 1 = ps_{i,j}, \forall (i,j) \in A \tag{14}$$

where,

$$ps_{i,j} = \begin{cases} 1, \text{a passive-type sensor is deployed on link } i\text{-}j \\ 0, \qquad\qquad\quad \text{otherwise} \end{cases};$$

N : node set;

A : link set;

R : a set of origin centroids, $R \subseteq N$;

S : a set of destination centroids, $S \subseteq N$.

The passive-type sensor location model is based on the link-node relationship, and uses double 0-1 binary variables to determine the nominal (minimal) required number of passive-type sensors. Eq. 9 is the objective function of the minimization problem, and is formulated as a linear program. Eq. 10 is the degree constraint required for full link flow observability. Eq. 11 is the "*reduplication constraint*". Based on the illustrated results shown in Table 2, to avoid selecting Link 3-6 twice, $b(3)_{3,6} + b(6)_{3,6} = 1 \geq 1$. Therefore one of the double 0-1 binary variables characterizing the same link in a neighboring node must be 1; for example, $b(3)_{3,6} = 0, b(6)_{3,6} = 1$ or $b(3)_{3,6} = 1, b(6)_{3,6} = 0$. Eqs. 12 and 13 are

the "*contradiction constraints*", and prevent bidirectional links from being selected as unequipped links, to maintain the flow conservation rule. Because the DC (Eq. 6) is derived from the flow conservation rule (Eq. 1), when a pair of bidirectional links is simultaneously selected as unequipped links, the flows on these bidirectional links cannot be inferred by using the flow conservation rule, and such a problem is defined as the "*contradiction problem*" in this research. Using the bidirectional Link 3-6 in Fig. 3 as an example, the double binary 0-1 variable results of selecting Link 3-6 as an unequipped link for Node 3, and Link 6-3 as an unequipped link for Node 6, are shown in Table 3; as indicated, the results remain within the degree constraints. Therefore, the flow across Link 3-6 for Node 3 is inferable based on the observed flow across Link 6-3, and the flow across Link 6-3 for Node 6 is inferable based on the observed flow across Link 3-6. This results in the contradiction problem; that is, when Links 3-6 and 6-3 are both selected as unequipped links, Link 3-6 requires the flow information of Link 6-3, and Link 6-3 also requires the flow information of Link 3-6. Consequently, flows on these two specific links are not inferable. Accordingly, we need the design of Eqs. 12 and 13 for double 0-1 binary variables in neighboring nodes, which can effectively resolve the contradiction problem.
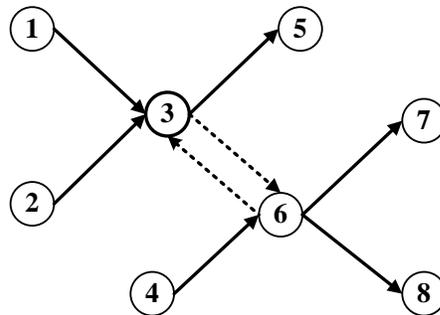
Fig. 3. Example network 3 for the contradiction problem.

Table 3. Unreasonable results for example network 3 for the contradiction problem.

Node #	Value of double 0-1 binary variable	Degree constraint
Node 3	$b(3)_{1,3}=1, b(3)_{2,3}=1,$ $b(3)_{3,5}=1, b(3)_{3,6}=0,$ $b(3)_{6,3}=1$	$DC_3=b(3)_{1,3}+b(3)_{2,3}+b(3)_{3,5}+b(3)_{3,6}+b(3)_{6,3}$ $=|ID_3|+|OD_3|-1=4$
Node 6	$b(6)_{3,6}=1, b(6)_{4,6}=1,$ $b(6)_{6,7}=1, b(6)_{6,8}=1,$ $b(6)_{6,3}=0$	$DC_6=b(6)_{3,6}+b(6)_{4,6}+b(6)_{6,7}+b(6)_{6,8}+b(6)_{6,3}$ $=|ID_6|+|OD_6|-1=4$

Eq. 14 is the "*identification equation*". According to Eq. 11, $1 \le b(i)_{i,j}+b(j)_{i,j} \le 2$. Eq. 14 yields a value of either 0 or 1; if 0, then the link is an unequipped link; if 1, then the link must be equipped with a passive-type sensor. Additionally, Eqs. 11 to 13 do not consider origin and destination nodes, it means that there is no $b(r)_{r,i}$ and $b(s)_{j,s}$. This research sets up the default values of $b(r)_{r,i}$ and $b(s)_{j,s}$ as 1 for the passive-type sensor deployment model.

2.3 *Active-type sensor location model*

The active-type sensors considered in this paper are license plate recognition (LPR) sensors. LPR sensors can collect information on link flows, vehicular trajectories, and route flows. The active-type sensor location model is based on the concept of path observability, and the path observability of a specific node is based on its incident links. Theoretically, the necessary condition for the full path observability of a given path is the number of adjacent links at a node minus 1. Using the simple network shown in Fig. 1 as an example, if active-type sensors are located on Links 1-3, 2-3, and 3-4, then the vehicular trajectories are 1-3, 2-3, 3-4, 1-3-4, and 2-3-4, where 1-3-4 and 2-3-4 are the complete paths, and 1-3, 2-3, and 3-4 are the incomplete paths. However, flows of 1-3-5

can be determined by comparing to flows of 1-3 and 1-3-4, and the difference between 2-3 and 2-3-4 represents 2-3-5. Because the condition of full path observability is equal to the degree constraints, the active-type sensor location model can be formulated similarly to the passive-type sensor location model, although the active-type sensor location model yields links with various weights; that is, it yields links with various priorities. Links connected to origin and destination nodes have a high selection priority, because their routes and/or O-D patterns can potentially be identified. Consequently, links connected to origin and destination nodes are assigned larger weights expressed in absolute values in this minimization program. The active-type sensor location model is formulated as follows:

$$\text{Min } \alpha \left(\sum_{(r,i) \in A} as_{r,i} + \sum_{(j,s) \in A} as_{j,s} \right) + \beta \left(\sum_{(i,j) \in A} as_{i,j} \right) \forall r \in R, s \in S \tag{15}$$

s.t.

Eqs. 10 through 13

$$b(i)_{i,j} + b(j)_{i,j} - 1 = as_{i,j}, \forall (i,j) \in A \tag{16}$$

where,

$$as_{i,j} = \begin{cases} 1, \text{an active-type sensor is deployed on Link } i\text{-}j \\ 0, \qquad\qquad\qquad \text{otherwise} \end{cases};$$

α, β : weight ($\alpha < \beta < 0$).

Eq. 15 is used to minimize the required number of active-type sensors, and links connecting to origin and/or destination nodes are selected first. Similarly, Eq. 16 is the identification equation for active-type sensors. In addition to Eq. 16, the active-type sensor location model includes Eqs. 10–13 as model constraints. The parameters, α and β, are the weights of centroid nodes connected links (hereinafter called *centroid links*) and intermediate nodes incident links (hereinafter called *intermediate links*),

respectively, and the absolute value of α must be greater than that of β to differentiate the relative importance of α and β in the minimization program. α and β, must both be less than 0 to prevent all solutions from being calculated as 0.

2.4   *The generalized sensor location model*

Because the passive-type and active-type sensor location models described are both subject to degree constraints, reduplication constraints, contradiction constraints, and identification constraints, these two heterogeneous sensor location models can be integrated to develop a generalized sensor location model. In this report, a generalized sensor location model is proposed based on heterogeneous sensor information under a budget constraint. The integrated sensor location model is formulated as follows:

$$\text{Min } \alpha\left(\sum_{(r,i)\in A} as_{r,i} + \sum_{(j,s)\in A} as_{j,s}\right) + \beta\left(\sum_{(i,j)\in A} as_{i,j}\right) + \gamma\left(\sum_{(i,j)\in A} ps_{i,j}\right) \forall r \in \boldsymbol{R}, s \in \boldsymbol{S} \qquad (17)$$

s.t.

Eqs. 10 through 13

$$b(i)_{i,j} + b(j)_{i,j} - 1 \geq as_{i,j} + ps_{i,j}, \forall (i,j) \in \boldsymbol{A} \qquad (18)$$

$$ca * \sum_{(i,j)\in A} as_{i,j} + cp * \sum_{(i,j)\in A} ps_{i,j} \leq TC \qquad (19)$$

where,

TC : total budget in traffic sensor deployment;
α, β, γ : weight ($\alpha < \beta < \gamma < 0$);
ca : the average cost of an active-type sensor;
cp : the average cost of an passive-type sensor.

Eq. 17 is a direct integration of Eqs. 9 and 15 based on weight. The parameters α, β, γ represent various weights. Because active-type sensors are able to collect greater amounts of flow distribution and flow information than passive-type sensors do, the absolute values of the parameters α, β are greater than that of γ. Because they capture different traffic and path patterns, deploying active-type sensors in centroid links must be prioritized; therefore, the absolute value of α is larger than that of β in minimization. To avoid the unreasonable solutions for the minimization problem, in which all solutions are 0, this study assumed that $\alpha<\beta<\gamma<0$; thus, the weights of passive- and active-type sensors are inversely proportional to their corresponding cost. A lower cost requires a higher absolute value weight ($|\gamma|<|\alpha|, |\gamma|<|\beta|$). The weights of active-type sensor links must reflect whether the link is connected to origin or destination nodes. A link connected to an origin or destination node is more important than a link connected to a noncentroid or intermediate node, but its maximal importance cannot be greater than two links connected to noncentroid nodes. Thus, its range is expressed as $|\beta|\leq|\alpha|<|2\beta|$. The generalized sensor location model also includes Eqs. 10–13. In contrast to the set covering rule, which requires known link-path incidence or O-D/path incidence matrices, the generalized sensor location model requires only a link-node incidence matrix and a 0-1 elements matrix, which is easily obtained according to a network's configuration. Eq. 18 is the identification equation for the generalized sensor location model. Eqs 14, 16, and 18 were designed to identify if a Link i-j is equipped with a sensor or without the need of deploying a sensor. Under no budget constraints, that is, no budget constraint is imposed on both Eqs. 14 and 16, this research defined double 0-1 binary variables, $b(i)_{i,j}$ and $b(j)_{i,j}$ to formulate the sensor deployment problem for the full observability on link flows. Such a variable design can avoid the reduplication and the contradiction problems. On the other hand, when a budget constraint is considered (e.g. Eq. 19), Eq. 18 is formulated as a greater than or equal relationship in the sense that the goal of full observability on link flows using both passive- and active-type sensor information might not be achieved due to insufficient monetary resources for sensor deployment. When the LFS of Eq. 18 is 1, there are two possible results for the RHS, listed below.

1) If the cost by adding a sensor on Link i-j does not violate the budget constraint, the RHS takes 1 (i.e. the equality relationship), meaning that Link i-j should be equipped with a sensor. Accordingly, the model will evaluate if Link i-j is equipped with a passive- or active-type sensor based on the relative contribution to the objective function.

2) If it violates the budget constraint, the RHS takes 0 (i.e. the inequality relationship). Under such a condition, Link i-j is not deployed a sensor since insufficient monetary sources are available for a sensor deployment.

On the other hand, when the LFS of Eq. 18 is 0, the RHS will be 0 as well (i.e. the equality relationship). Link i-j is not equipped a sensor. In summary, when the LFS of Eq. 18 is 1, the RHS will be 0 or 1; if the LFS of Eq. 18 is 0, the RHS will be 0. Such a "greater than or equal" relationship for Eq. 18 is specifically formulated to realistically reflect practical limitations on budget constraints for sensor deployment. Finally, Eq. 19 is the budget constraint.

Note that in the objective function (Eq. 17), passive-type sensors do not distinguish origin and destination links (i.e. centroid links). The main purpose of deploying passive-type sensors is to collect link flow information for full link flow observability purpose. On the other hand, installing active-type sensors at different types of links has different implications. Specifically, active-type sensors deployed at centroid links are more valuable for path flow or O-D demand estimation than deploying at intermediate links (i.e. noncentroid links). If active-type sensors are sequentially installed at an origin, a destination, and an intermediate link, it can possibly identify path (flow) patterns by matching path information of these three active-type sensors. Such collected link flow and partial path (flow) patterns are crucial inputs to a network O-D demand estimation model.

CHAPTER 3. LINK-BASED NETWORK O-D DEMANDS ESTIMATION

Chapter 3 studies the network O-D demands estimation problem by using the flow information given by different sensor sources. This chapter discusses the application of the link flow and partial path flow information respectively given by passive- and active-type sensors. A nonlinear program for the network O-D demands estimation problem is formulated and solution algorithms are described. In section 3.1 the nonlinear O-D demand estimation model is described. In section 3.2 solution algorithms for both the sensor location problem and network O-D demands estimation problem are given.

3.1   *The O-D demands estimation model*

Based on the generalized sensor location model, a subset of path flows can be obtained. Past studies assumed that link-path incidence matrices are known; however, link-path incidence matrices are practically difficult to obtain because they require path enumeration. To relax this unreasonable assumption, this study developed an O-D demand estimation model in which path information is not required. Nguyen (1977) proposed a non-proportional traffic equilibrium assignment model based on the node-arc formulation. Nguyen's model was modified by further incorporating the link flow conservation rule into an easily obtained link-node incidence matrix to develop the O-D demand estimation model presented in this study, which is formulated as follows:

$$\text{Min} \sum_{(i,j)\in A} (\overline{y}_{i,j} - \hat{y}_{i,j})^2 + \sum_{r}^{|R|} \sum_{s}^{|S|} (\hat{t}^{rs} - \sum_{p=1}^{|P|} \overline{f}_p^{rs})^2 \tag{20}$$

s.t.

$$\hat{Y} \cdot LN = 0 \tag{21}$$

$$\sum_{r\in R}^{|R|} \hat{t}^{rs} = \sum_{j\in N} \hat{y}_{j,s}, s \in S \tag{22}$$

$$\sum_{s\in S}^{|S|} \hat{t}^{rs} = \sum_{i\in N} \hat{y}_{r,i}, r \in R \tag{23}$$

$$\hat{t}^{rs} \geq \sum_{p=1}^{|P|} \overline{f}_p^{rs} \tag{24}$$

$$\hat{t}^{rs} \geq 0 \tag{25}$$

where,

$\hat{Y}$: a vector of estimated link flows;

$\overline{y}_{i,j}$: observed Link i-j flows;

$\hat{y}_{i,j}$: estimated Link i-j flows;

$\overline{f}_p^{rs}$: the path flows between pair r-s using the pth path collected from active-ty

$\hat{t}^{rs}$: an unknown O-D flow between pair r-s;

LN : link-node incidence matrix.

The O-D demand estimation model is a nonlinear minimized squared program. Eq. 20 minimizes two terms. The first term is the difference between the measured and estimated link traffic flows. The second term is the difference between the observed and the estimated O-D flows. Observed O-D flows are partially obtained by aggregating the path flows associated with specific O-D pairs provided by active-type sensors. Eq. 21 is

the link flow conservation constraint determined according to the link-node incidence matrix, which does not include origin or destination nodes. Eq. 22 yields the total origin flow to a specific destination node, which is equal to the total link flows connected to this specific destination node. Eq. 23 yields the total destination flow from a specific origin node, which is equal to the total link flows connected to this specific origin node. Eq. 24 is the inequality constraint, and implies that true O-D flows are greater than or equal to the summation of observed (partial) path flows. Eq. 25 is the nonnegative constraint. The decision variables in this estimator are the unknown O-D flows; this model does not assume known path and/or prior O-D information, route choice probabilities, or turning proportions, and is thus feasible for practical application.

This research uses observed link flow information and partial path (flow) patterns as input data for O-D demand estimation without the need of a known link-path incidence matrix and/or the prior O-D information to search for feasible O-D demand solutions. The solution for the proposed O-D demand estimation model may not be unique due to the underdetermined system and the limited flow information, but it can find out possible and reasonable O-D demand solutions with the assistance of active-type sensor information. For a nonlinear optimization problem, it can be relaxed by first (second) order Taylor series expansion, and this relaxation can be further transformed into a linear program with a set of variables. This set of variables can be regarded as the descent of original variables, and the optimal solution is based on sequentially solving the descent until it is optimal. If the descent is not efficiently converged, one can consider the radius of the descent; the optimal criterion is based on that the radius is within a pre-specified acceptable tolerance. For the network O-D demand estimation problem, despite it is difficult to obtain a unique solution on O-D demand estimates, the second best approach to solve the O-D demand estimation problem under a certain level of tolerance or error is available.

3.2　*Solution Algorithms*

The proposed generalized sensor location model is used to determine the minimal required number of passive-type and active-type sensors, considering degree constraints, contradiction constraints, budget constraints, and the identification equation. Because the decision variable is a 0-1 integer, the objective function and constraints are linear integer equations, and the model can be determined using a 0-1 integer program. The mathematical optimization tool LINGO 11 (LINDO Systems Inc., Chicago, IL) was used to solve the generalized sensor location model.

The O-D demand estimation problem was formulated using a nonlinear least squares (NSL) program with linear constraints, and it was solved using the nonlinear solver of LINGO 11. The first-order derivate was calculated in the nonlinear solver using backward analytical derivatives, and successive linear programming (SLP) directions were used to determine directions. First-order Taylor series expansion, a linear approximation method, was adopted for computations to reduce iteration time.

CHAPTER 4. CASE STUDIES

Chapter 4 evaluates the proposed generalized sensor location model for network O-D demands estimation purposes through numerical tests using two networks: a fishbone network (Hu et al., 2009) and a simplified real road network. Section 4.1 conducts the numerical test in light of the fishbone network. Section 4.2 evaluates the proposed models using a simplified real road network at NCKU. Section 4.3 provides a discussion on the link-based and path-based flow estimation approaches for the NSLP.

4.1 *The fishbone network*

4.1.1 *Fishbone network without budget constraints*

The generalized sensor location model was evaluated without budget constraints, with the additional constraint that active-type sensors could not exceed two-thirds of the total number of deployed sensors. This additional constraint ensured that all deployed sensors were not active-type sensors or passive-type sensors. The results indicated that 12 sensors (4 active-type sensors and 8 passive-type sensors) were required. This number conformed to the upper-bound ratio proved by Ng (2012), $(m\text{-}n)/m$, where m is the number of links, and n is the number of noncentroid nodes ([18-6]/18). The sensor location configuration is shown in Fig. 4; the red dotted arrows denote active-type sensors, and the blue dotted arrows denote passive-type sensors.
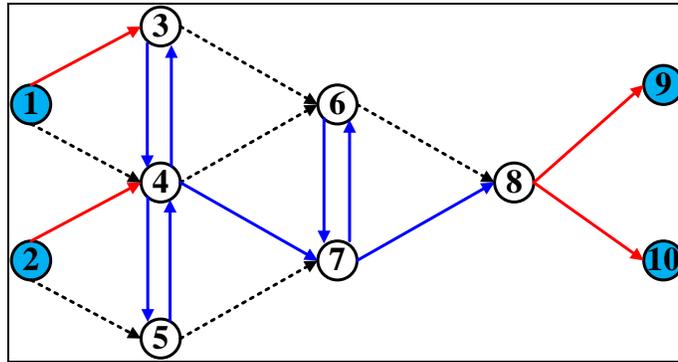
Legend:

⬤ (shaded node): O-D node

◯ (unshaded node): intermediate node

⇢ (dotted arrow): unequipped link

→ (red solid arrow): active-type sensor equipped link

→ (blue solid arrow): passive-type sensor equipped link

Fig. 4. Result of the NSLP for the fishbone network without budget constraint.

This sensor location configuration achieved the full observability of link flows, and each unequipped link flow could be inferred. The results of two test scenarios are shown in Table 4; the first test scenario involved executing O-D demand estimation by incorporating flow information provided by passive-type sensors, and the second test scenario involved executing O-D demand estimation by incorporating both link flow information provided by passive-type sensors and partial path flow information provided by active-type sensors. These two test scenarios used the same number of sensors, but the sensors were deployed at different links. Based on the information provided by the four active-type sensors, partial traffic patterns were identified. These reduced the average MAPE from 76.87% to 6.98%, a 70% improvement of the average MAPE for the estimation of network O-D demands.

Table 4. Results for the fishbone network without budget constraint.

Data source(s)	O-D	T_{1-9}	T_{1-10}	T_{2-9}	T_{2-10}	Average MAPE for O-D flow estimation	Average MAPE for link flow inference
	True volume	98	223	237	387	--	--
Passive-type sensor	Estimated volume	203	422	436	85	--	--
	MAPE	107.14%	89.24%	83.97%	27.13%	76.87%	0%
Passive- & active-type sensors	Estimated volume	85	236	250	374	--	--
	MAPE	13.27%	5.83%	5.49%	3.36%	6.98%	0%

4.1.2   *Fishbone network under budget constraints*

To evaluate the model under budget constraints, this study assumed that the costs of passive-type and active-type sensors were 20 and 80, respectively, and the total budget was 470. The results indicated that four passive-type sensors and four active-type sensors could be installed at various strategic links (see Fig. 5), for a total of eight sensors. Because of the budget constraint, information on certain unequipped links could only be partially inferred based on the observed link flows. The average MAPE for the link flow inference was 10.89%, indicating that, under the budget constraint, certain link flows were not inferable. Based on the numerical results of the O-D demand estimation shown in Table 5, even under a budget constraint, the O-D demand estimation model still provided error-free results, even on unequipped link flows and O-D flows in small networks with fewer links and paths. The flow estimation results were not substantially affected by the route choices of users.
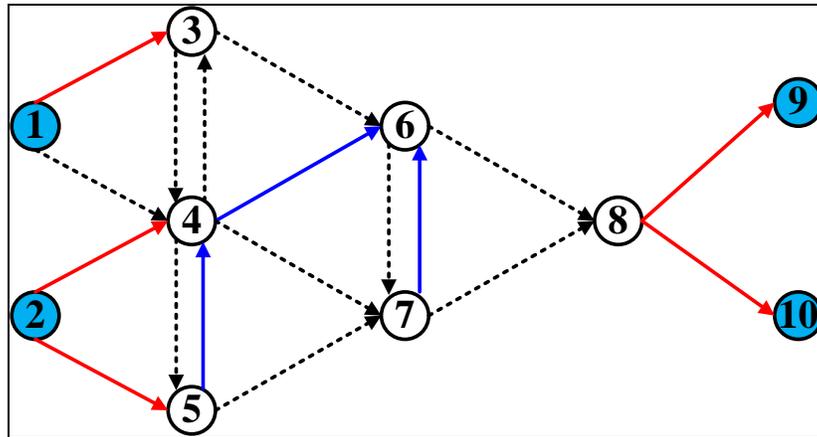
Fig. 5. Results of the NSLP in the fishbone network with budget constraint.

Table 5. Results for the fishbone network with budget constraint.

Data source(s)	O-D	T_{1-9}	T_{1-10}	T_{2-9}	T_{2-10}	Average MAPE for O-D flow estimation	Average MAPE for link flow inference
	True volume	98	223	237	387	--	--
Passive- & active-type sensors	Estimated volume	98	223	237	387	--	--
	MAPE	0.0%	0.0%	0.0%	0.0%	0.0%	10.9%

As shown in Tables 4 and 5, the results indicated that full observability of link flows may result in the partial observability of O-D demand flows. In addition, partial observability of link flows may also result in full observability in O-D demand estimation.

4.2   *The NCKU network*

The NCKU network (Fig. 6) is located in the eastern district of Tainan City in Southern Taiwan, bordering Tainan Station (Node 32 in Fig. 6). The NCKU network consists of seven main campuses of NCKU, including the Li-Hsing (Node 1), Chien-Kuo (Node 2), Ching-Yeh (Node 3), Kuang-Fu (Node 4), Cheng-Kung (Node 5), Tzu-Chiang

(Nodes 6 and 8), and Sheng-Li (Node 7) campuses. Nodes 1 to 8 are the O-D on-campus nodes, and Nodes 9 to 14 are the O-D off-campus nodes. The NCKU network has three urban arterials connected to the National Sun Yat-Sen freeway system: the first is Dongfeng Rd., represented by Nodes 15, 16, 17, 18, 19, and 20; the second is Xiaodong Rd., represented by Nodes 10, 22, 23, 24, 25, 26, 27, 28, and 14; and the third is Dongning Rd., represented by Nodes 32, 33, 34, 35, 36, 37, 38, and 39. Four urban streets cross the main campuses: the first is University Rd., represented by Nodes 32, 33, 34, 35, 36, 37, 38, and 39; the second is Shengli Rd., represented by Nodes 16, 21, 24, 29, 35, 40, 43, and 45; the third is Changrong Rd., represented by Nodes 18, 26, 30, 37, 41, 44, and 47; and the fourth is Linsen Rd., represented by Nodes 20, 28, 31, 39, 42, and 48.

Legends:

 (red shaded node): on-campus O-D node

○ (blue shaded node): off-campus O-D node

○ (unshaded node): intermediate node

→ (black solid arrow): directed link or arc
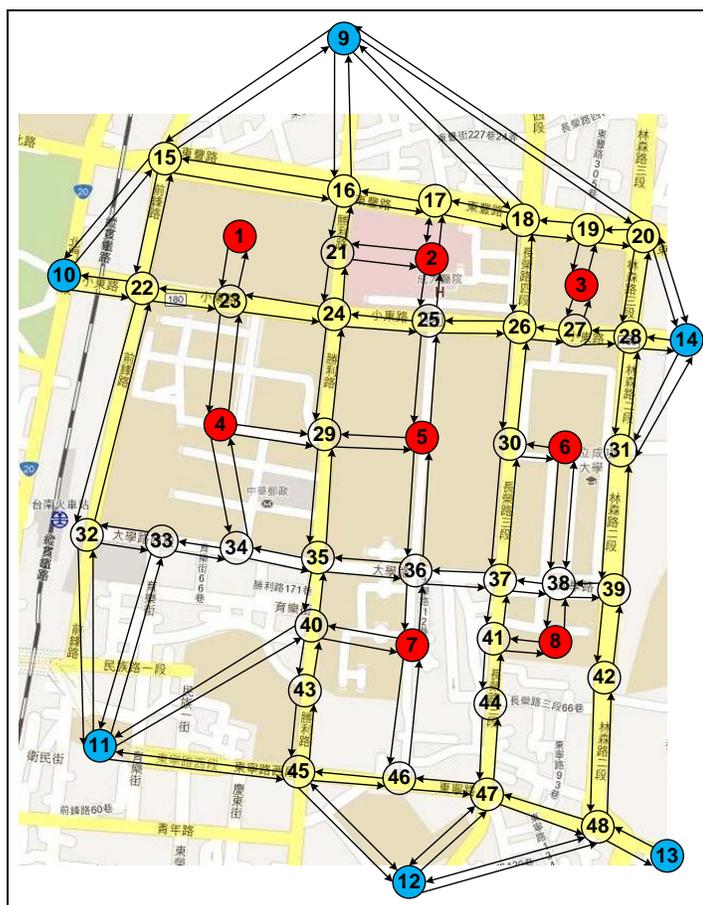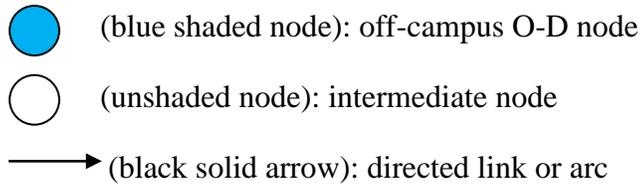
Fig. 6. NCKU network.

The true O-D dataset was obtained from the Tainan City home survey (Yen et al., 2000) conducted by COMDYCS Technology Consultants, Inc. This data set was used to evaluate the performance of the proposed O-D demand estimation model, and the generalized sensor location model with and without budget constraints.
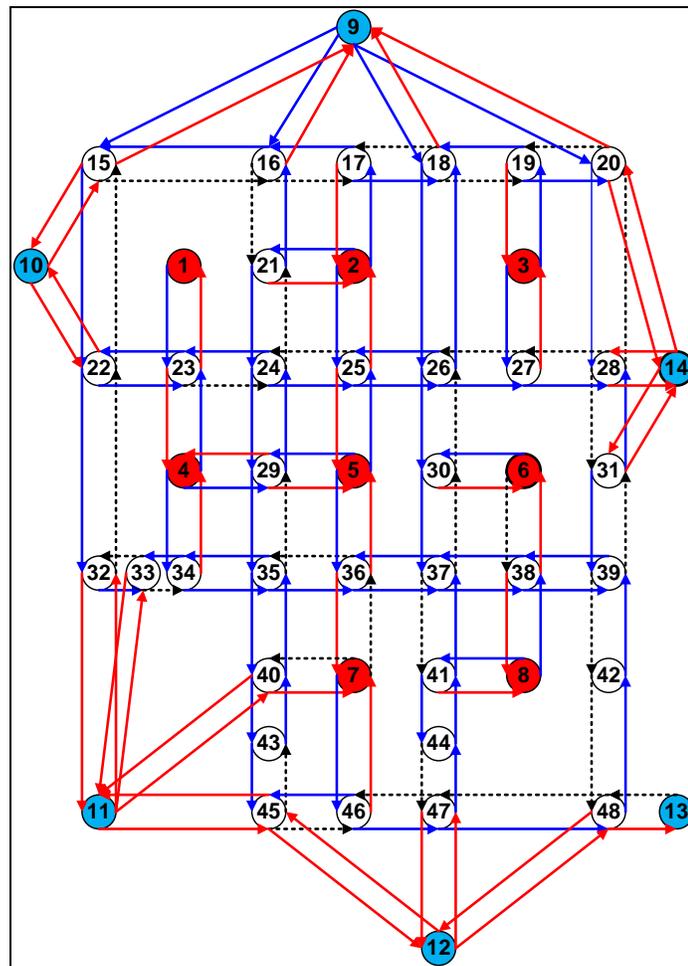
Legends:

(red shaded node): on-campus O-D node

(blue shaded node): off-campus O-D node

(unshaded node): intermediate node

(black dotted arrow): unequipped link

(red solid arrow): active-type sensor equipped link

(blue solid arrow): passive-type sensor equipped link

Fig. 7. Results of the NSLP in the NCKU network without budget constraints.

4.2.1   *The NCKU network without budget constraints*

The results indicated that the NCKU network required 48 active-type sensors and 72 passive-type sensors. The total number of deployed sensors was 120; this conformed to the upper-bound ratio proved by Ng (2012) ([154-34]/154). The sensor location configuration is shown in Fig. 7; solid red arrows represent links monitored by deployed active-type sensors, and dotted blue arrows represent links monitored by passive-type sensors.

The O-D demand estimation results for the NCKU network, without budget constraints, are shown in Fig. 8 in terms of the number of O-D pairs corresponding to different MAPE ranges. The results of two different scenarios are shown: the first scenario involved using information provided by passive-type sensors, and the second scenario involved using information provided by both passive- and active-type sensors. Without budget constraints, both scenarios achieved full observability of link flow inference (that is, the average MAPEs for link flow inference in the two test scenarios were 0%) when using the maximal number of traffic sensors. Using only passive-type sensor information to estimate network O-D demands provided biased O-D demand estimates for certain O-D pairs. The average MAPE for estimating O-D flows by using

only passive-type sensor information was 45.21%. Incorporating partial path information provided by active-type sensors into the proposed O-D demand estimation model substantially reduced the MAPEs for certain O-D pairs, and the average MAPE for estimating O-D flows by using information provided by both passive- and active-type sensors was 21.46%. Checking the variations of MAPEs, there are 12 O-D pairs whose MAPEs are larger than 100% using only passive-type sensor information (T_{2-7}, T_{3-7}, T_{4-6}, T_{5-2}, T_{5-3}, T_{8-6}, T_{9-1}, T_{9-4}, T_{9-10}, T_{13-12}, T_{13-15}, T_{14-8}). When the partial path flow information given by active-type sensors is applied to the NCKU network, there are 7 O-D pairs whose MAPEs are reduced from >100% to <10% (T_{2-7}, T_{3-7}, T_{4-6}, T_{5-2}, T_{5-3}, T_{8-6}, T_{14-8}), 3 O-D pairs' MAPEs are reduced to <75% (T_{9-1}, T_{9-4}, T_{13-15}), and 2 O-D pairs' MAPEs are reduced to <100% (T_{9-10}, T_{13-12}). The MAPEs for the remaining O-D pairs still keep a similar trend.

When active-type sensors are incorporated into the O-D demand estimation problem, the MAPEs for some O-D pairs are reduced and some are increased. To fairly evaluate of the performance of the proposed O-D demand estimation model, this research defines a relative assessment index, called Reduced MAPE (RMAPE), defined in Eq. 27.

$$RMAPE^{r\text{-}s} = oMAPE^{r\text{-}s} - wMAPE^{r\text{-}s} \tag{27}$$

where,

    $RMAPE^{rs}$: reduced MAPE for O-D pair r - s;

    $oMAPE^{rs}$: MAPE for O-D pair r - s without active-tpye sensor information;

    $wMAPE^{rs}$: MAPE for O-D pair r - s with active-tpye sensor information.

For the NCKU network without budget constraint case, 12 O-D pairs had positive RMAPEs among 141 O-D pairs. The maximal RMAPE is 700%, and the minimal RMAPE is -84.85%. There were only five O-D pairs with negative RMAPEs in the totally investigated 141 O-D pairs. Among the total 141 O-D pairs, the total RMAPE is

3348.91%; the average of RMAPE is 23.75%. The RMAPE evaluation results indicated that averagely the MAPE for each O-D pair can be reduced by 23.75% by incorporating active-type sensor information.
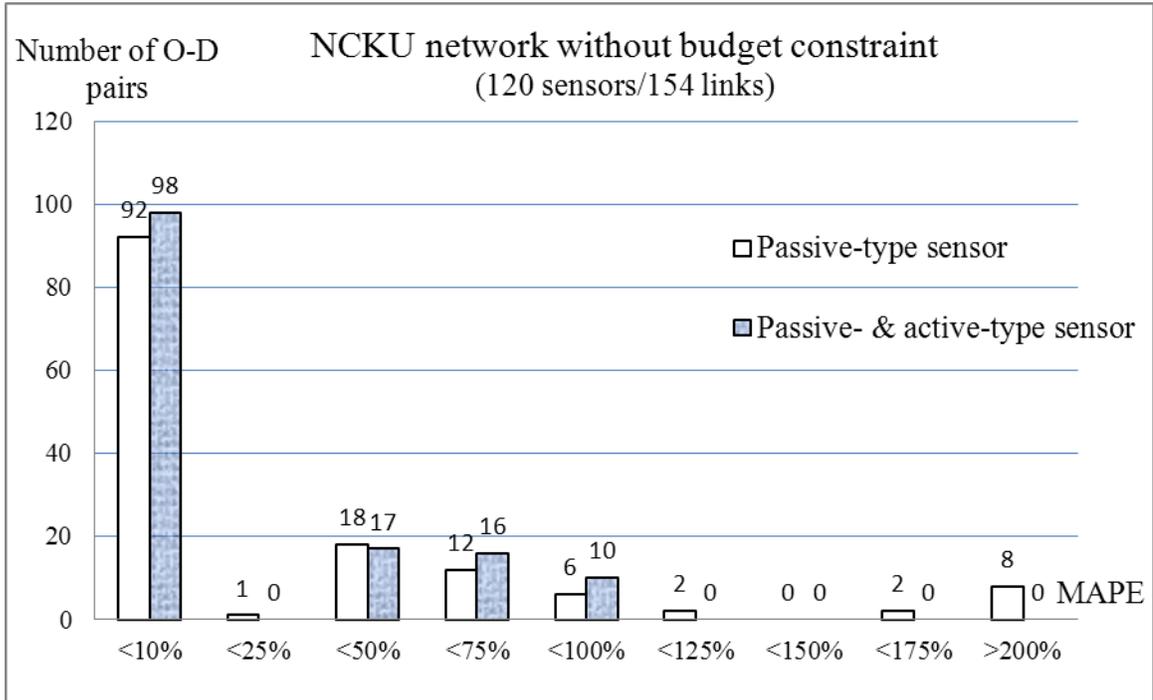
Fig. 8. Results for the NCKU network without budget constraint.

4.2.2 *The NCKU network under budget constraints*

To evaluate the model in the NCKU network under budget constraints, this study assumed weights of $\alpha = -8, \beta = -4, \gamma = -2$ in Eq. 17, that the costs of passive-type and active-type sensors were 20 and 80, respectively, and that the total budget was 3000. The results indicated that 69 sensors were required: 27 passive-type and 42 active-type sensors (see Fig. 9). Only partial link flows across the links without sensor monitoring capability can be inferred based on the observed link flows because of the budget constraints. The average MAPE for the link flow inference was 62.13% because of the low sensor deployment rate (69/154 = 49%). The O-D demand estimation results, shown

in Fig. 10, indicated that the performance of the O-D demand estimation model was satisfactory, and that the results improved with the incorporation of partial path flow information provided by active-type sensors. The average MAPE for O-D flow estimation using information provided by both passive- and active-type sensors was 20.67%.
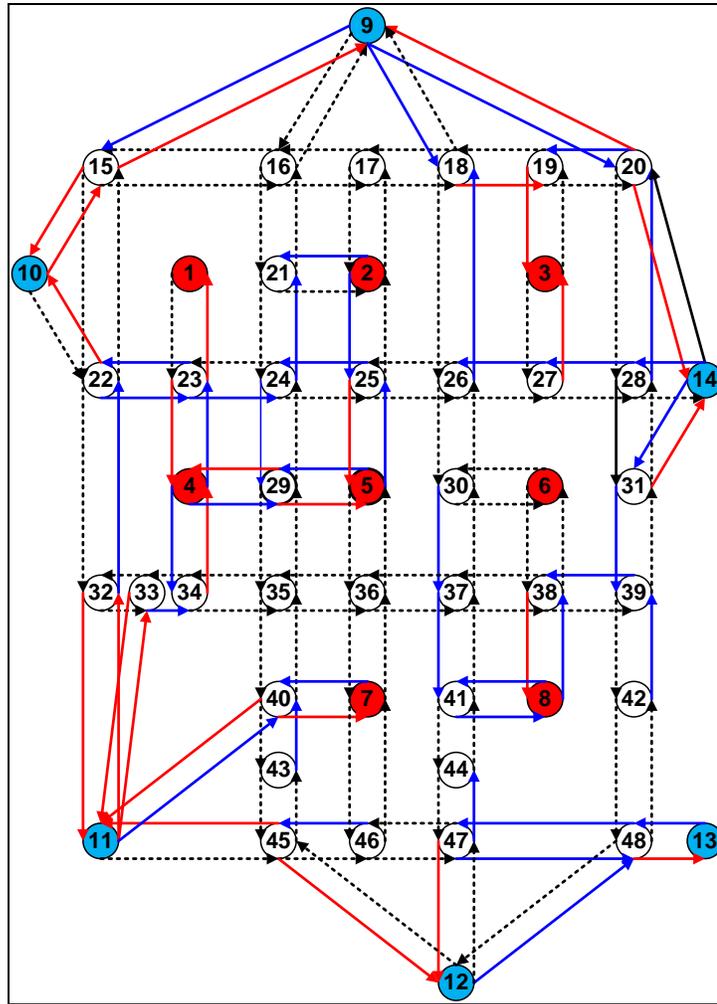
Fig. 9. Results of the NSLP in the NCKU network with budget constraint.

The results under budget constraints indicated that partial information provided by passive-type and active-type sensors was sufficient to obtain satisfactory O-D demand estimates. This implies that additional link flow inputs in the O-D demand estimation

model do not necessarily result in improved O-D demand estimation performance, because link flow information may be either linearly dependent on each other or redundant. However, partial link and path flow information provided by passive- and active-type sensors deployed at certain critical links could be used to obtain satisfactory solutions to the O-D demand estimation problem. Theoretically, the quality of network O-D demand estimation is highly dependent on the quality of link and/or path flow information. Because of budget constraints, active-type and passive-type sensors can be partially deployed only at certain strategic links in a target network, potentially resulting in biased link flow estimates. However, regardless of inaccurate link flow inferences, incorporating information on partial path flow patterns provided by deployed active-type sensors improves O-D demand estimates. Path flow and/or traffic pattern information provided by active-type sensors is highly useful in the estimation of network O-D demands. For the variations of MAPE ranges, it also gets the similar result to the no budget condition. The O-D pairs with >100% MAPE can be reduced. More O-D pairs can obtain <10% MAPE.

For the NCKU network under a budget constraint case, there were 34 O-D pairs with positive RMAPEs among the total 141 O-D pairs. The maximal RMAPE is 700%. The minimal RMAPE is -99.68%. In addition, there were 23 O-D pairs with negative RMAPEs in the total 141 O-D pairs. Among the total 141 O-D pairs, the total RMAPE is 3459.90%, the average RMAPE is 24.54%. The RMAPE evaluation results indicated that averagely the MAPE for each O-D pair can be reduced by 23.54% by incorporating active-type sensor information.
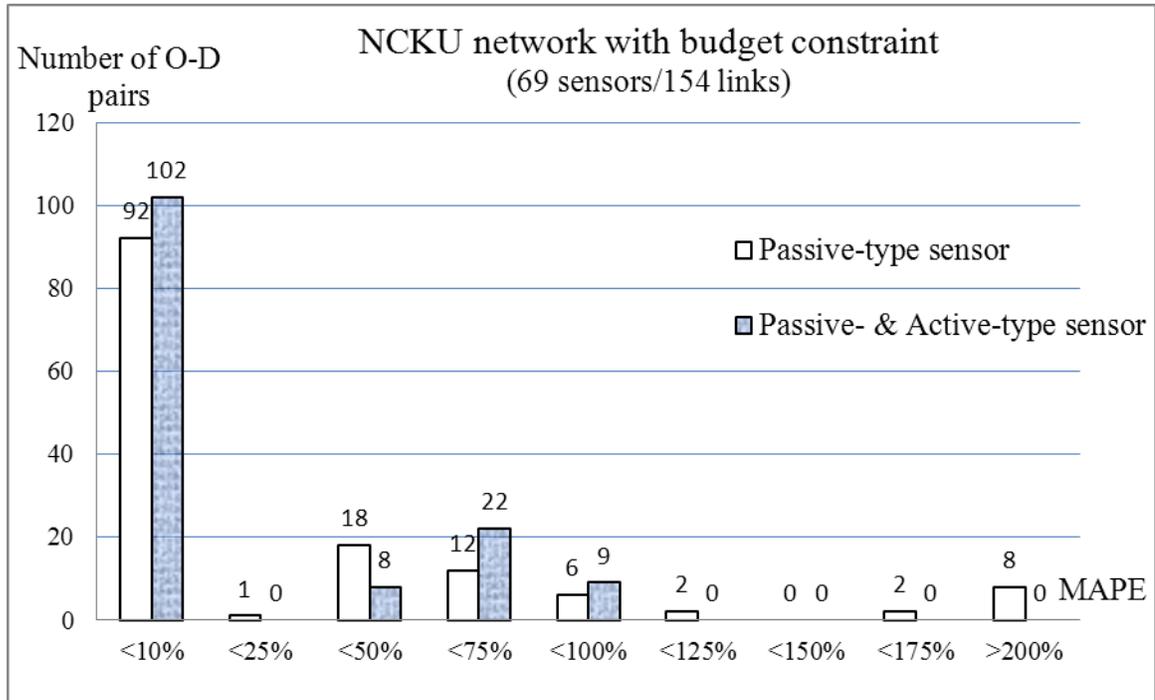
Fig. 10. Results for the NCKU network with budget constraint.

4.3   *Discussion: path-based versus link-based approaches*

Most studies on the estimation of network O-D demands have assumed the availability of known link-path incidence or O-D/path/link incidence matrices. Based on this specific matrix, a set of O-D and/or paths has respective relationships with the intermediate links, and results of O-D/path flows estimation are highly dependent on the quality and quantity of observed link flows. However, in practice, link-path incidence and O-D/path/link incidence matrices are not easily obtained because of the path enumeration problem. The proposed link-based O-D demand estimation model estimates path flows based on a non-proportional traffic assignment principle, and relaxes this unreasonable assumption; moreover, this model is able to obtain satisfactory estimates on O-D flows. The estimated O-D demands are constrained by links connected to origin and destination nodes or centroid links (see Eqs. 22 and 23), and intermediate nodes are constrained by the flow conservation rule (see Eq. 21). In this proposed model, the O-D demand

estimation result is highly dependent on links connected to centroid nodes. Because of budget constraints, certain link flows could not be satisfactorily inferred. Nevertheless, the results of the estimated O-D demands can be improved by appropriately installing active-type sensors at certain strategic links to obtain partial path pattern and/or flow information.

CHAPTER 5. CONCLUSIONS

This chapter summarizes the research, highlights its contributions, and proposes directions for future research.

5.1   *Summary*

This study addresses the two primary objectives:

1. Propose an effective generalized sensor location model for sensor location flow-observability and sensor location flow-estimation problems.

2. Give an assumption-free, link-based network O-D demands estimation formulation taking advantage of the flow information given by different sources of sensors.

5.2   *Future research directions*

This study solved the sensor location and network O-D demand estimation problems in two steps. The first step focused on the generalized sensor location model, and the second step focused on the O-D demand estimation model. These two steps were independent of each other. We propose two future research directions. First, bi-level programming, in which the upper level is a heterogeneous traffic sensor location model and the lower level is an O-D demand estimation model, can be studied, and based on the correlation between the upper and lower levels. Second, the heterogeneous traffic sensor location model can be integrated with the network O-D demand estimation model in a single-step framework. A bi-level, integrated model framework would be more straightforward and useful in practical applications.

REFERENCES

Hu, S. R., Peeta, S., Chu, C. H. (2009). Identification of vehicle sensor locations for link-based network traffic applications. Transportation Research Part B, 43(8-9), p. 873-894.

Ng, M. (2012). Synergistic sensor location for link flow inference without path enumeration: A node-based approach. Transportation Research Part B, 46(6), p. 781-788.

Nguyen, S. (1977). Estimating an OD Matrix from Network Data: A Network Equilibrium Approach. Centre for Research on Transportation, University of Montreal Publication 60.

Yen, Y. M., Hsiao, W. C., Hsu, H. H., Lee, L. J., Yeah, M. H. (2000). Household Travel Characteristics Survey in Tainan Metropolitan. Institute of Transportation, Ministry of Transportation and Communications. Serial No. 90-37-5211, Taipei, Taiwan.