Inferring Network Origin-Destination Matrices Using Partial Link Traffic Flow Information

Han-Tsung LIOU
Graduate Research Assistant
Department of Transportation and Communication Management Science
National Cheng Kung University
No. 1, University Road, Tainan City 70101, Taiwan
Phone: 886-6-2757575 ext. 53271 ext. 2090;
E-mail: iroya.liou@gmail.com

Shou-Ren HU
Assistant Professor
Department of Transportation and Communication Management Science
National Cheng Kung University
No. 1, University Road, Tainan City 70101, Taiwan
Phone: 886-6-2757575 ext. 53203;
Fax: 886-6-2753882;
E-mail: shouren@mail.ncku.edu.tw

Abstract: Inferring trip origin-destination (O-D) matrix in a vehicular network using partial link flow information is a critical issue for transportation planning and traffic management. However, due to a budgetary constraint, the minimum subset of links to be equipped with vehicle detectors (VDs) and their installation locations need to be determined to provide a desirable O-D matrix estimate. In the present research, a linear algebra-based method was developed to deal with the network location problem. Given a network structure represented by its link-path incidence matrix, some crucial links were strategically identified by a basis link model (Hu *et al*., 2009a, b) and the collected link flow information were further incorporated into an O-D matrix estimation model based on link flow conservation rule without any unreasonable assumptions, such as known prior O-D information and/or users' route choice probability. Numerical results based on a simplified real highway network indicated that the O-D matrix estimates given by the basis link model combining with the proposed O-D estimation model are generally satisfactory.

Key Words: *trip origin-destination, vehicle detector, basis link, link-path incidence matrix*

1. INTRODUCTION

Trip origin-destination (O-D) pattern is one of several crucial elements that affect traffic situation and trip distribution in a vehicular network. An O-D matrix in a transportation and/or highway network represents the travel patterns/demands on a zonal network at a specific time period. These matrices describe both spatial issue of trip distribution and temporal issue of traffic dispersion of the corresponding trip demands. For transportation planning purposes, an O-D demand matrix significantly affects the decisions associated with travel direction, path selection, and trip length. For traffic control and/or management usage, time-dependent O-D demands are essential inputs to the determination of a desirable traffic control scheme. Therefore, trip O-D pattern in a vehicular network is one of the fundamental inputs to the transportation planning, traffic engineering, network design, and traffic management areas.

In the area of O-D matrix/demand estimation, traditional data collection means, such as roadside interview, postcard survey, and/or license-plate investigation are becoming infeasible due to the problems associated with time consuming, very costly, labor intensive and

sampling error. On the contrary, as the rapid developments of Intelligent Transportation Systems (ITS) and traffic detection technologies, inferring network trip O-D demands through the observed link traffic flow information becomes economically feasible. However, it is difficult to deploy a full scale of VDs at all links for the highway management agency due to budgetary constraints. Therefore, how to strategically install VDs at some crucial/partial links that provide necessary flow information for network O-D demand estimation purpose becomes a critical issue that is worthy of further investigation. The main purpose of this research is to infer trip origin-destination (O-D) matrices in a vehicular network using partial link traffic flow information. To deal with this problem, a basis link model (Hu *et al.*, 2009a, b) was adopted to obtain the necessary link flow information. These collected link flows were further incorporated into a network O-D demand estimation model. Numerical analysis based on a simplified real highway network indicated that the proposed O-D demand estimation model generally provides satisfactory results.

The remainder of this paper is organized as follows. Section2 describes the O-D estimation and traffic counting location problems and provides reviews of the relevant literatures. The model formulations regarding VDs deployment strategy, O-D demand estimation, and the corresponding solution algorithms are developed and presented in Section 3. Section 4 presents the experimental design and model evaluation results based on the numerical analysis. Finally, in Section 5, findings are summarized and future research directions are suggested.

2. PROBLEM DESCRIPTION AND LITERATURE REVIEW

In the past two decades, most O-D matrix estimation researches generally employ synthetic techniques to obtain different variants of O-D matrix estimates. In general, these synthetic approaches can be categorized into two families. The first family methods aim to obtain O-D demands by using simple aggregate behavioral rules. The logic behind this family is based on home or roadside interview, or postcard survey to estimate O-D demands at the specific origins and destinations and then adopts simple behavioral assumptions/rules for estimating O-D matrices. The methods belong to the other family use traffic flow data as input to the O-D estimation models. The main advantage of traffic flow data is that they are easily obtained by vehicle detectors, providing approaches of this family with an economically feasible solution to the O-D demand estimation problem.

In general, O-D demands can be classified into static and dynamic cases, depending on the temporal issue of traffic dispersion and their applications. Specifically, static O-D demand data either collected by traditional means or inferred by link traffic flows cannot reflect temporal variations of traffic dispersion. They are usually applied to transportation planning and facility location problems. Dynamic O-D demand data, on the other hand, are the short-term trip distributions between specific O-D pairs; they are important inputs to traffic control and/or management systems. In the literature, the major methods for estimating O-D demand using a time series of traffic flow data include Linear Least Squares (LLS) and Nonlinear Least Squares (NLLS) (e.g., Cascetta and Nguyen, 1988; Cremer and Keller, 1987; Davis, 1993; Nihan and Davis, 1987; Zhou and Mahmassani, 2006), Maximum Likelihood (ML) (e.g., Nihan and Hamed, 1992, Spiess, 1987; Cascetta *et al.*, 1993), Entropy Maximizing (EM), (e.g., Lam and Lo, 1991; Van Zuylen and Willumsen, 1980), Bayesian inference (e.g., Maher, 1983; Cascetta and Nguyen, 1988; Castillo *et al.*, 2008c), Kalman Filtering (KF) (e.g., Ashok and Ben-Akiva, 1993; Chang and Wu, 1994; Ashok and Ben-Akiva, 2002), and Artificial Neural Network (ANN) (e.g. Suzuki *et al.*, 2001). Besides,

some extended researches estimate network O-D demands using vehicle probe by statistical sampling theory (e.g., Hellinga and Van Aerde, 1994), automatic vehicle identification (AVI) using market penetration approach (e.g., Eisenman and List, 2003; Dixon and Rilett, 2005) or license plate scanning (e.g. Castillo *et al*., 2008d).

Various methods in the literature for O-D demand estimation have different degrees of capabilities to provide O-D demand estimates, but most of the research work assume that link traffic flow information is either available for each link of the observed network (e.g., Bianco, *et al*., 2001; Cascetta and Nguyen, 1988; Cremer 1987; Lam and Lo, 1991; Nihan and Davis, 1987; Nihan and Hamed, 1992; Nguyen, 1984, 1977; Spiess, 1987; Van Zuylen and Willumsen, 1980) or can be collected at some specific locations, such as screen link (Wu and Chang, 1996) or partial link flow information (Sherali *et al*., 2003). Moreover, some of the research work assume that known prior information on O-D demands or users' route choice probabilities. Such assumptions might be unrealistic for practical applications. Since route choice probabilities which are the traffic distribution outcomes of traffic assignment given some prior O-D information, which are generally not known in advance. Therefore, we aim to develop an O-D demand estimation model based on a non-proportional traffic equilibrium assignment model (e.g. Nguyen, 1977), and a mathematic programming model formulated by Leblanc and Farhangian (1982) which is suitable for dealing with the O-D demand estimation problem.

Since the quantity and quality of collected link traffic flow information significantly affect the estimation accuracy and reliability of O-D estimates, there is a trade-off between the accuracy of O-D demand estimates and the cost associated with the deployment of a VD system. In view of a limited resource constraint, there is an urgent need of a VD deployment strategy to provide accurate O-D demand estimates. To determine the strategic VD installation locations for the estimation of network O-D matrices, most studies in the literature deal with the detector location problem as a sub-problem within a broader network O-D estimation problem, and numerous VD deployment rules/principles have been proposed. For instance, graph theory (Binanco *et al*., 2001), schema theory (Horiguchi *et al*., 2001), linear independent (LID) links (Hu and Wang, 2008; Hu *et al*., 2008), basis link method (Hu and Peeta, 2008; Hu *et al*., 2009a, b). Among them, Yang and Zhou (1998) proposed four sensor location rules: O-D cover rules, maximal flow fraction rules, maximal flow-intercepting rule, and link independence rule, which are most comprehensive principles for various network location problems. Some studies address the locations of VDs as an observability problem (Casotillo *et al*. 2007, 2008a, 2008b) and propose null space-based method (Casotillo *et al*. 2007), algebraic and topological methods (Casotillo *et al*., 2008a, 2008b) to determine the optimal number and locations for VDs deployment. However, state-of-the-art methods concerning the VD deployment problem for network O-D matrix estimation generally assume that prior O-D information and/or users' route choice probabilities are available (e.g., Yang and Zhou, 1998; Casotillo *et al*. 2007, 2008a, 2008b, 2008c). To avoid the unrealistic assumptions stated above, this research aims to obtain partial link traffic flow information under a strategic VD deployment plan and conduct O-D estimation based on a non-assignment based O-D demand estimation model.

The VD deployment problem in a general network is essentially a long-term transportation planning and network location problem, since it is unusual and difficult to change the installation locations in a short period of time once a VD deployment plan has been implemented. Thus it is reasonable to evaluate a VD deployment strategy by considering a traffic flow pattern under a user's equilibrium assignment assumption, such as UE or

stochastic UE (SUE) traffic assignment principle. The purpose of the present research is therefore to solve a VD deployment problem given a network structure represented by its link-path incidence matrix, and further tackle the O-D demand estimation problem accordingly under a flow conservation rule without involving any unreasonable assumptions on known prior O-D information and/or users' route choice probability.

3. MODEL FORMULATION

To deal with the VD deployment problem for network O-D demand estimation, a linear algebra-based method proposed by Hu *et al*. (2009a, b) is developed. Given some known vehicle trajectories obtained by some reasonable traffic assignment rules (e.g., the SUE), this research develops an O-D demand estimation method using partial link traffic flow information. To determine the VD deployment strategy, a general network is regarded as a system of linear equations of vehicle trajectories. Based on this linear equation system, a link-path incidence matrix is constructed. Sine there is no need to collect full link traffic flow information due to redundant information on the link-path incidence matrix, we may employ only some key (column) vectors of the matrix to infer an entire vector space. This research introduces the concept of "basis" that contains the properties of linear independence (LID) and span to deal with the VD deployment problem. Based on the concept of basis in a vector space, it can efficiently determine the critical links on a link-path incidence matrix and the number and installation locations of required VDs to provide a certain degree of accuracy on an O-D matrix. The network O-D demand estimation problem is solved by a pseudo-inverse matrix method. The O-D demand estimates are obtained given minimum number of VDs and meanwhile avoiding the unrealistic assumptions on known prior O-D demand information and/or route choice probabilities. The entire model framework includes three components:1) a link-path incidence matrix based on known vehicle trajectories, 2) a VD deployment model, 3) a trip O-D estimation model, and they are discussed as follows.

3.1 Link-path incidence matrix

When conducting the O-D estimation problem, it is necessary to consider the optimal number and installation locations of VDs due to budget constraints. Before determining an optimal VD deployment plan, this research introduces the link-path incidence matrix proposed by Cho (1991, 1992) to describe a general network by 0-1 binary variables. Let L be an $(k \times m)$ link-path incidence matrix and it can be written as follows:

$$\mathbf{L} = \{l_{ij}\}, \forall i \in \{1,2,3,...,k\},\ j \in \{1,2,3,...,m\} \tag{1}$$

where, $l_{ij} = 1$, it means that link a_j belongs to path p_i; and $l_{ij} = 0$, otherwise.

Equation (1) indicates that L is an $(k \times m)$ link-path incidence matrix including m links and p paths. Clearly, a link-path incidence matrix is (0-1) matrix representing the trajectories of the paths traversing different links in a general network.

3.2 VD deployment model

To determine an optimal VD deployment plan, a concept of basis in linear algebra (Friedberg, *et al*., 2003) is proposed. Based on the concepts of linear algebra, a matrix contains two vector sets: basis vector and non-basis vector; and non-basis vector is generally linearly dependent (LD) on basis vector. Basis contains the properties of linear independence (LID) and span. Figure 1 describes the properties of basis.

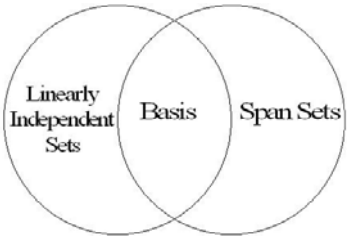

Figure 1 Properties of basis

Based on a set of basis, it is possible to span a vector space by vector adding or scalar multiplication. For a link-path incidence matrix, it also can be classified into basis links and non-basis links (Hu *et al*., 2009a, b). This research introduces a solution algorithm of reduced row echelon form (RREF) to distinguish those LID columns or links in a link-path incidence matrix, and further find out the basis links. A matrix is said to be in RREF if it satisfies the following properties (Friedberg, *et al*., 2003):

1). Any row containing a nonzero entry proceeds any row in which all the entries are zero (if any).

2). The first nonzero entry in each row is the only nonzero entry in this column.

3). The first nonzero entry in each in each row is 1 and it appears in a column to the right of the leading 1 in any proceeding row.

In general, the RREF can be easily calculated by the Gaussian-Jordan elimination method based on elementary row operation. According to the RREF algorithm, it easily obtains a set of basis from a link-path incidence matrix. By implementing a Gaussian-based elimination method, we can obtain a reduced row echelon form (RREF) of the original matrix and a column containing leading 1 of a RREF is an LID. In a RREF, there will be a set of basis links among the LID links. Based on the concept of basis, this research solves the network VD location problem by the RREF algorithm.

To illustrate the concept of basis of a matrix identified by the RREF algorithm, take a small network for example, a network is consisted of one origin and two destinations, and four paths and nine links (see Figure 2).

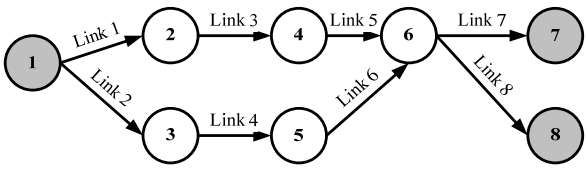

Figure 2 Small sample network

The used paths in the small sample network are as follows. For O-D pair 1-7, P1: 1→2→4→6→7 and P2: 1→3→5→6→7; and for O-D pair 1-8, P3: 1→2→4→6→8 and P4: 1→3→5→6→8. The original link-path incidence matrix is therefore constructed and shown in Table 1, and its corresponding RREF is shown in Table 2.

Table 1 Original link-path incidence matrix.

number	1	2	3	4	5	6	7	8
link / path	1--2	1--3	2--4	3--5	4--6	5--6	6--7	6--8
1 (1→2→4→6→7)	1	0	1	0	1	0	1	0
2 (1→3→5→6→7)	0	1	0	1	0	1	1	0
3 (1→2→4→6→8)	1	0	1	0	1	0	0	1

4 (1→3→5→6→8)	0	1	0	1	0	1	0	1

Table 2 RREF of the original link-path incidence matrix.

number	1	2	3	4	5	6	7	8
link / path	1--2	1--3	2--4	3--5	4--6	5--6	6--7	6--8
1 (1→2→4→6→7)	1	0	1	0	1	0	0	1
2 (1→3→5→6→7)	0	1	0	1	0	1	0	1
3 (1→2→4→6→8)	0	0	0	0	0	0	1	-1
4 (1→3→5→6→8)	0	0	0	0	0	0	0	0

It is shown in Table 2 that links 1, 2, 3, 4, 5, 6, and 7 are LID links. But links 3 and 5 are completely linearly dependent on link 1, and links 4 and 6 are also fully captured by link 2. After deleting some links with the same entries, one can obtain a set of basis links in the RREF, and the result is shown in Table 3.

Table 3 A set of basis link in the RREF

number	1	2	3	4	5	6	7	8
link / path	1--2	1--3	2--4	3--5	4--6	5--6	6--7	6--8
1 (1→2→4→6→7)	1	0	1	0	1	0	0	1
2 (1→3→5→6→7)	0	1	0	1	0	1	0	1
3 (1→2→4→6→8)	0	0	0	0	0	0	1	-1
4 (1→3→5→6→8)	0	0	0	0	0	0	0	0

Based on contents of Table 3, a set of basis links are links 1, 2 and 7. The solution of a set of basis in a matrix may not be unique, such as links 1, 2 and 7 or links 1, 4 and 7 or links 2, 3 and 7 or links 5, 6, and 7 are all sets of basis. Nevertheless, no matter which set of the basis links, it can span the original (link-path incidence) matrix. In theory, the different sets of basis are *row equivalent*; the different sets of basis inputting to an O-D demand estimation model should obtain the same result. There may be no optimal solution set of the basis links, however we may give different weighting factor for each link, such safety ranking, benefit/cost (B/C) ratio and network connectivity, etc, in order to obtain a specific optimal solution of the basis links set. In this research, we assume that all of links are equally weighting and the goal is to find a set of desirable solution for O-D demand estimation purpose. This research proposes the basis links are the strategic locations for VD deployment since they can fully describe the entire flow pattern of the observed network. Accordingly, the rank of matrix, such as (Friedberg *et al*., 2003), is the required VDs to be deployed and we can use this set of basis link traffic flow information to infer O-D demands, which will be described below.

3.3 Trip O-D demand model

This research proposes a flow conversation rule to conduct O-D demand estimation and introduces the pseudo-inverse matrix method by incorporating the concept of 2-norm (Friedberg *et al*., 2003; Penny and Lindfield, 2000) to obtain a set of O-D demand estimates. The O-D demand estimation model under the basis link traffic flow information is formulated as:

$$\sum_{j=1}^{r} \hat{f}_i \times l_{ij}^b = x_j^b \ \forall i = 1,2,...,k \ \forall j = 1,2,...,r \tag{2}$$

where,

$\hat{f}_i$ is the estimation flow of the p_i path;

l_{ij}^b is the basis coefficient of link-path incidence matrix, L;

x_{ij}^b is the flow of the basis link;

r is the rank of link-path incidence matrix, L.

Equation (2) is named the flow conservation equation. If one transposes equation (2), it can be written as follows:

$$(\mathbf{L}^B)^{\mathrm{T}}\hat{\mathbf{F}}^T = (\mathbf{X}^B)^{\mathrm{T}} \tag{3}$$

where,

$\mathbf{L}^B$ is an $(k \times r)$ basis of link-path incidence matrix;

$\mathbf{X}^B$ is an $(1 \times r)$ basis link flow information;

$\hat{\mathbf{F}}$ is an $(1 \times k)$ estimation path flow information.

Based on equation (3), it can be written in a system of linear equations, the solution of this system depends on the rank of the link-path incidence matrix, such that full rank gets unique solution, rank-deficiency gets no solution or multiple solutions. This research follows the pseudo-inverse matrix method to conduct the rank-deficiency problem.

3.4 Solution Algorithm

Based on the link flow conservation equation, it obviously indicates that the system of linear equations is an under-determined system since the number of paths is usually greater than the number of links. For overcoming this problem, this research introduces the pseudo-inverse matrix method which is able to effectively deal with the over-, under-determined systems and/or rank deficiency problems to obtain a minimum norm least squares solution (MNLSS). Firstly, the 2-norm for the vector $\hat{F}^T$ is defined by:

$$\left\|\hat{\mathbf{F}}^T\right\|_2 = (\left|\hat{f}_1\right|^2 + \left|\hat{f}_2\right|^2 + \ldots + \left|\hat{f}_k\right|^2)^{1/2} \tag{4}$$

Equation (4) is referred to the Euclidean norm of the vector of $\hat{F}^T$. Euclidean norm is also called the length of the vector. In the following, this research follows the pseudo-inverse matrix method and incorporates the concept of 2-norm to solve the unknown variable vector in equation (3). The MNLSS of equation (3) is shown in equation (5).

$$\hat{\mathbf{F}}^T = (\mathbf{L}^B)^{T^+}(\mathbf{X}^B)^T + (\mathbf{I} - (\mathbf{L}^B)^{T^+}(\mathbf{L}^B)^T)\mathbf{Y} \tag{5}$$

where,

$(\mathbf{L}^B)^{T^+}$ is the pseudo-inverse of $(\mathbf{L}^B)^T$;

$\mathbf{Y}$ is an arbitrary vector;

$\mathbf{I}$ is an identity matrix.

Equation (5) is the general solution of $\hat{\mathbf{F}}^T$, which is the derivate form the pseudo-inverse matrix, (Ben-Israel and Greville, 2003; Cho, 1991, 1992; Friedberg, *et al*., 2003; Penny and Lindfield, 2000). The approximation of the MNLSS in equation (5) can be computed in two states (two-stage minimization problem):

State 1:

$$\min_F \left\|(\mathbf{X}^B)^{\mathrm{T}} - (\mathbf{L}^B)^{\mathrm{T}}\hat{\mathbf{F}}^T\right\|_2 \tag{6}$$

State 2:

$$\min_F \left\|\hat{\mathbf{F}}^T \text{ among all solution of stage 1}\right\|_2 \tag{7}$$

Equation (6) tries to minimize the error between the estimates and observations, and equation

(7) prevents the solutions from being the null space. Besides, based on equations (6) and (7), a minimum norm least square solution can be obtained while satisfying the system of equations. A pseudo-inverse matrix generally gives three types of solutions under different conditions. Equation (8) is for row full rank, Equation (9) is for column full rank, and Equation (10) is for rank deficiency and it is called Singular Value Decomposition (SVD).

$$(\mathbf{L}^B)^{T^+} = [(\mathbf{L}^B)(\mathbf{L}^B)^T]^{-1}(\mathbf{L}^B) \tag{8}$$

$$(\mathbf{L}^B)^{T^+} = (\mathbf{L}^B)[(\mathbf{L}^B)^T(\mathbf{L}^B)]^{-1} \tag{9}$$

$$(\mathbf{L}^B)^{T^+} = (\mathbf{V}^B)[(\mathbf{S}^B)^T(\mathbf{S}^B)]^{-1}(\mathbf{S}^B)^T(\mathbf{U}^B)^T \tag{10}$$

where,

$\mathbf{V}^B$ is an orthogonal $(k \times k)$ matrix;

$\mathbf{U}^B$ is an orthogonal $(r \times r)$ matrix;

$\mathbf{S}^B$ is an $(r \times k)$ singular value matrix.

According to the static known vehicle trajectories, this research addresses a concept of basis as the strategy of VD deployment and uses the pseudo-inverse matrix method incorporating a 2-norm concept to conduct the link flow conservation equations in order to obtain a set of O-D demand estimates.

4. NUMERICAL TESTS

In this section, some network scenarios are proposed to evaluate the VD deployment configuration and O-D estimation performance. For the known vehicle trajectories, this research can construct a corresponding link-path incidence matrix by *EXCEL 2003*. According to the procedure of RREF algorithm, it can easily find out the optimal location and number of VDs and further use the pseudo-inverse matrix method to solve the O-D estimation with considering about the link conservation rule. These procedures are calculated by *MATLAB 7.5*. Since the vehicle trajectories condition may be altered by other traffic assignment principles or user choice probability, this research will give the various link-path incident matrices for an existence network and evaluate the results for the O-D estimation under an optimum VD deployment configuration. This research adopts the mean absolute percent error (MAPE) as the evaluation criterion, and it is defined as follows:

$$\text{MAPE} = [\frac{100}{K}\sum_{K=1}^{K}\left|\frac{e_k - o_k}{o_k}\right|] \tag{11}$$

where,

e_k is the *k-th* O-D estimate under basis link information;

o_k is the *k-th* O-D demands;

K is the total number of O-D pairs.

There are many evaluation criteria for measuring error, such as mean squared error (MSE), mean absolute percent error (MAPE), root mean squared error (RMSE), root relative mean square error (RRMSE)…etc. There are some reasons for using MAPE as the evaluation criterion in this research, 1) it is very simple to calculate, 2) The MAPE not only considers the absolute value which can give the positive penalty for under- or over-estimation but also considers the relative measure, 3) it has a basic reference index (Lewis, 1982) to evaluate the error (see Table 4). Typical MAPE values for industrial and business data and their

interpretation are shown in Table 4.

Table 4 Interpretation of typical MAPE values

MAPE (%)	Interpretation
< 10%	Excellent
10%--20%	Good
20%--50%	Reasonable
> 50%	Bad

Source: Lewis (1982).

4.1 Scenario-1 (The Various Link-path Incidence Matrices)

In scenario-1, this research will test the different link-path incidence matrices for the same network and evaluate their optimal VD deployment configuration and O-D estimation. The network is shown in Figure 3.

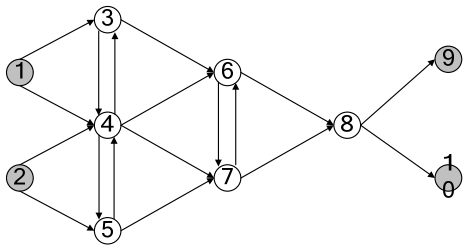

Figure 3 Test network-1

The test network-1 network is consisted of 10 nodes, 18 links, 4 O-D pairs (T_{1-9}, T_{1-10}, T_{2-9}, T_{2-10}), and the O-D demands are 98, 223, 237, 387, respectively. For test network-1, the research gives three kinds of vehicle trajectories conditions, it means that there are three link-path incidence matrices (22, 26, 29 paths, respectively), there are shown in Table 5.

Table 5 Three kinds of vehicle trajectories conditions

Vehicle Trajectories / O-D Pairs	(1)	(2)	(3)
1-9	1-3-6-8-9, 1-4-6-8-9 1-4-7-8-9, 1-3-6-7-8-9 1-4-5-7-8-9	1-3-6-8-9, 1-4-6-8-9 1-4-7-8-9, 1-3-4-7-8-9 1-3-6-7-8-9	1-3-4-6-9, 1-4-6-8-9 1-4-7-8-9, 1-3-4-7-8-9 1-4-6-7-8-9, 1-4-7-6-8-9
1-10	1-3-6-8-10, 1-4-6-8-10 1-4-7-8-10, 1-3-4-7-8-10 1-3-6-7-8-10	1-3-6-8-10, 1-4-6-8-10 1-4-7-8-10, 1-3-4-7-8-10 1-3-6-7-8-10	1-3-6-8-10, 1-4-6-8-10 1-4-7-8-10, 1-3-4-6-8-10 1-3-4-7-8-10, 1-4-3-6-8-10 1-4-6-7-8-10, 1-4-7-6-8-10
2-9	2-4-6-8-9, 2-4-7-8-9 2-5-7-8-9, 2-5-4-3-6-8-9 2-4-5-7-6-8-9, 2-5-4-7-8-9	2-4-6-8-9, 2-4-7-8-9 2-5-7-8-9, 2-4-3-6-8-9 2-4-5-7-8-9, 2-5-4-6-8-9 2-5-4-7-8-9,2-5-7-6-8-9	2-4-6-8-9, 2-5-7-8-9 2-4-3-6-8-9, 2-4-6-7-8-9 2-5-4-7-8-9, 2-5-7-6-8-9 2-5-4-7-6-8-9
2-10	2-4-6-8-10, 2-4-7-8-10 2-5-7-8-10, 2-5-4-7-8-10 2-5-7-6-8-10, 2-5-4-3-6-8-10	2-4-6-8-10, 2-4-7-8-10 2-5-7-8-10, 2-4-3-6-8-10 2-4-5-7-8-10, 2-5-4-6-8-10 2-5-4-7-8-10, 2-5-7-6-8-10	2-4-6-8-10, 2-4-6-8-10 2-4-7-8-10, 2-5-7-8-10 2-4-5-7-8-10, 2-5-4-7-8-10 2-5-7-6-8-10, 2-5-4-7-6-8-10

After applying the RREF algorithm and the pseudo-inverse matrix method into three kinks of link-path incidence matrices of test network-1, the optimal VD deployment configuration is shown in Figure 4.

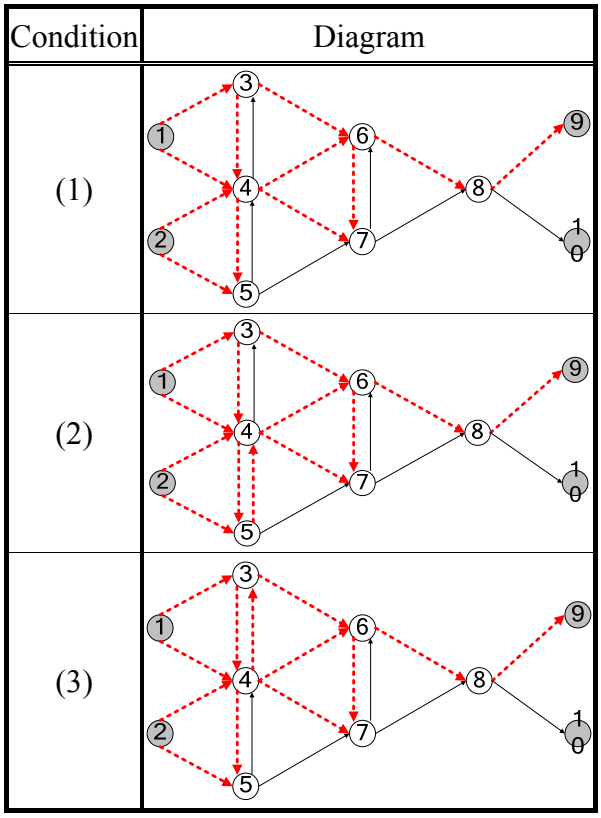

Figure 4 The optimal VD locations in test network-1
(the boldface dotted lines represent the basis links equipped with VDs)

Based on Figure 4, there are different vehicle trajectories conditions, they may result in the similar the optimal VD locations, the optimal number of VDs are 12, 13, 13, respectively, and they are also similar. The O-D estimations in test network-1 are resulted in Table 6,

Table 6 The results of O-D estimation in test network-1

Condition		(1)				(2)				(3)			
O-D Pairs	O-D Demands	Full Info.	MAPE	Basis Info.	MAPE	Full Info.	MAPE	Basis Info.	MAPE	Full Info.	MAPE	Basis Info.	MAPE
1-9	98	109.29	11.52%	109.29	11.52%	110.64	12.90%	110.64	12.90%	93.01	5.09%	93.01	5.09%
1-10	223	211.71	5.06%	211.71	5.06%	197.18	11.58%	197.18	11.58%	214.06	4.01%	214.06	4.01%
2-9	237	225.71	4.76%	225.71	4.76%	231.04	2.52%	231.04	2.52%	232.67	1.83%	232.67	1.83%
2-10	387	398.29	2.92%	398.29	2.92%	406.14	4.95%	406.14	4.95%	405.25	4.72%	405.25	4.72%
Avg. MAPE			6.07%		6.07%		7.99%		7.99%		3.91%		3.91%
Number of basis link				12				13				13	
% of links be equipped with VDs		100%		67%		100%		72%		100%		72%	

From Table 6, it indicates that full link information and basis link information can get the same results of O-D estimation. Approximately, 60%-70% of links equipped with VDs in a

network will be helpful to infer O-D information. It means that it is no necessary to collect all of links information, and only basis links information will be available for O-D estimation. Each individual and average MAPEs value of O-D estimation is below 10%, it means the results of O-D estimation are excellent and/or good in term of interpretation of typical MAPE values.

4.2 Scenario-2

In secenario-2, this research based on a large network which was presented in Hu *et al*. (2008), this large network contains 52 nodes, 154 links, 15 O-D pairs (T_{1-9}, T_{1-21}, T_{3-5}, T_{4-28}, T_{6-33}, T_{6-27}, T_{8-52}, T_{10-22}, T_{21-35}, T_{32-41}, T_{37-46}, T_{41-6}, T_{46-21}, T_{50-6}, T_{52-5}) and it was evaluated in two kinds of vehicle trajectories, 57 and 121 paths, respectively. The optimal VD locations are shown in Figure 5 and Figure 6, respectively and the results of O-D estimations are resulted in Table 7,

Table 7 The Results of O-D Estimation in Test Network-2

Condition		(1) (57 paths-containing)				(2) (121 paths-containing)			
O-D Pairs	O-D Demands	Full Info.	MAPE	Basis Info.	MAPE	Full Info.	MAPE	Basis Info.	MAPE
1-9	1,980	1991.35	0.57%	1991.35	0.57%	1949.39	1.55%	1949.39	1.55%
1-21	1,746	1700.74	2.59%	1700.74	2.59%	1792.68	2.67%	1792.68	2.67%
3-5	2,220	2249.01	1.31%	2249.01	1.31%	2212.9	0.32%	2212.9	0.32%
4-28	1,217	1329.23	9.22%	1329.23	9.22%	1215.75	0.10%	1215.75	0.10%
6-33	1,391	1468.05	5.54%	1468.05	5.54%	1368.47	1.62%	1368.47	1.62%
6-27	1,538	1425.11	7.34%	1425.11	7.34%	1547.13	0.59%	1547.13	0.59%
8-52	1,771	1739.69	1.77%	1739.69	1.77%	1764.73	0.35%	1764.73	0.35%
10-22	1,956	1834.56	6.21%	1834.56	6.21%	1952.38	0.18%	1952.38	0.18%
21-35	1,270	1271.31	0.10%	1271.31	0.10%	1262.66	0.58%	1262.66	0.58%
32-41	1,678	1693.75	0.94%	1693.75	0.94%	1680.89	0.17%	1680.89	0.17%
37-46	1,340	1343.22	0.24%	1343.22	0.24%	1347.5	0.56%	1347.5	0.56%
41-6	1,825	1833.56	0.47%	1833.56	0.47%	1822.62	0.13%	1822.62	0.13%
46-21	1,505	1464.05	2.72%	1464.05	2.72%	1514.2	0.61%	1514.2	0.61%
50-6	1,648	1645.81	0.13%	1645.81	0.13%	1646.95	0.06%	1646.95	0.06%
52-5	1,748	1843.55	5.47%	1843.55	5.47%	1754.75	0.39%	1754.75	0.39%
Avg. MAPE			2.96%		2.96%		0.66%		0.66%
Number of basis link				42				83	
% of links be equipped with VDs		100%		27.27%		100%		53.90%	

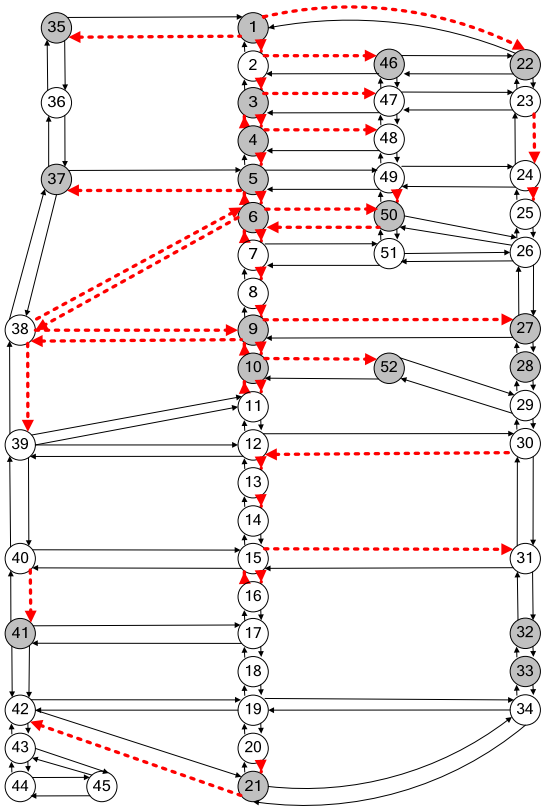

Figure 5 The optimal VD locations in test network-2 (57 paths-containing)
(the boldface dotted lines represent the basis links equipped with VDs)

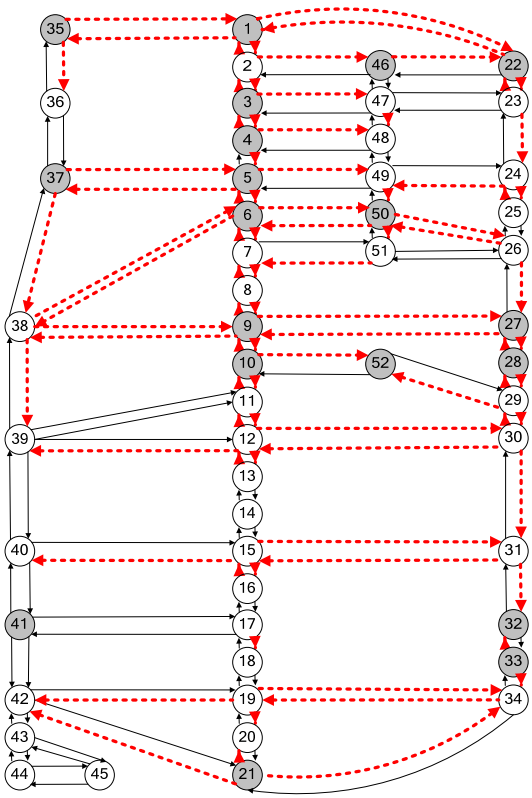

Figure 6 The optimal VD locations in test network-2 (121 paths-containing)
(the boldface dotted lines represent the basis links equipped with VDs)

In scenario-2, they also get the same results of the O-D estimations under full or basis link information. Form Table 7, it also can get the good estimation results and it reasonably indicates that the more number of paths needs more basis link information to infer O-D information.

4.3 Insights

According to the results of experimental design and numerical analysis, some findings and their implications are summarized below:

1. The pseudo-inverse matrix method generally provides satistactory O-D demand estimates, no matter under full link flow information or partial (basis) link flow information.
2. For a network represented by its link-path incidence matrix, the more number of used paths, the more basis link information needed to infer an O-D demand estimate, unless this network is in a saturated condition; even if more paths are considered, it will not obviously influence the number of basis links.
3. Considering the basis links as the optimal VD deployment scheme, it obviously indicates that if there is only some critical link information considered, it can be very useful to infer O-D demand estimates within a certain degree of estimation accuracy.
4. Another unique aspect of this research is that we avoid users' route choice and/or assignment probabilities to infer an O-D demand estimate by the link flow conversation rule.

5. CONCLUSIONS AND RECOMMENDATIONS

The RREF algorithm applied to identify the basis links presented in this research offers a better opportunity to strategically deploy VDs in a general network, with the goal of collecting critical link flow information to infer network O-D demands. One of the unique features of the proposed model is that it is able to provide relatively good network O-D demand estimates under a limited number of link traffic flow information without full-scale VD deployment, which, in turns, saves the initial cost of a traffic monitoring and/or detection system. In addition, the proposed O-D estimation model framework also avoids the unreasonable assumptions on known prior O-D information, turning proportions, and/or route choice probabilities, which are difficult to collect or unavailable in practice.

According to the numerical analysis results, effective network O-D demand information can be obtained by deploying VDs in a strategic manner. The number of VDs deployed is generally less than the total number of links in a general network (30% - 70%). Moreover, the number of required VDs conforms to the MAPE performance criteria in different network scales or the number of O-D pairs. In addition, the number of required VDs also gradually increases as the investigated network becomes larger. The pseudo-inverse matrix method is able to provide the same result based on either full or basis link information and obtains satistactory estimation results on an O-D matrix in light of a minimum norm least squares solution. Namely, it is not necessary to collect entire full link information based on a full-scale VD deployment; inferring full information and O-D demands by partial link flow information is not only feasible but also desirable.

Finally, this research focuses only on a static condition for the optimal deployment of link-based VDs. For RREF, it is based on the linear algebra and it gives us a most reduced matrix form, its solution will be equal to other linear algebraic-based method, such as pivot. If the number of VDs needs to be further reduced, the link-based detector information is not enough; it necessarily combine with other path-based detector information, such as automatic

vehicle identification (AVI), license plate scanning and probe data, etc. These advanced vehicle detection equipments may provide a better result of deployment of link-based vehicle detectors and O-D demand estimation. Nevertheless, the cost associated with path-based detectors will be greater than cost of the link-based detectors, it is not desirable in real work. The O-D demand estimates given by the proposed methods in light of basis link information are unknown constants. In reality, even for the static case, they are essentially random variables. Therefore, future study may develop a stochastic mathematical programming model and the according solution algorithms to deal with the stochastic network location problem.

REFERENCES

Ashok, K. and Ben-Akiva, M.E. (1993) Dynamic origin-destination matrix estimation and prediction for real-time traffic management systems, in C. F. Daganzo, editor, **International Symposium on Transportation and Traffic Theory**, Elsevier Science Publishing Company, Inc., 465-484.

Ashok, K. and Ben-Akiva, M.E. (2002) Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows, **Transportation Science, Vol. 36, No. 2**, 184-198.

Bianco, L., Confessore, G. and Reverberi, P. (2001) A network based model for traffic sensor location with implications on O/D matrix estimates, **Transportation Science, Vol. 35, No. 1**, 50-60.

Ben-Israel, A. and Greville, T. N. E. (2003) **Generalized inverses: theory and applications**, Springer, New York.

Cascetta, E., Inaudi, D. and Marquis, G. (1993) Dynamic estimators of origin-destination matrices using traffic counts, **Transportation Science, Vol. 27, No. 4**, 363-373.

Cascetta, E. and Nguyen, S. (1988) A unified framework for estimating or updating Origin-Destination matrices from traffic counts, **Transportation Research-B, Vol. 22B, No. 6**, 437-455.

Castillo, E., Conejo, A.J., Pruneda, R.E. and Solares, C. (2007) Observability in linear systems of equations and inequalities. Applications, **Computers and Operation Research, Vol. 34, No. 6**, 1708-1720.

Castillo, E., Conejo, A.J., Menéndez, J.M. and Jiménez, P. (2008a) The observability problem in traffic network models, **Computer-Aided Civil and Infrastructure Engineering, Vol. 23, No. 3**, 208-222.

Castillo, E., Jiménez, P., Menéndez, J.M. and Conejo, A.J. (2008b) The observability problem in traffic models: algebraic and topological methods, **IEEE Transactions on Intelligent Transportation Systems, Vol. 9, No. 2**, 275-287.

Castillo, E., Menéndez, J.M. and Sanchez-Cambronero, S. (2008c) Traffic estimation and optimal counting location with path enumeration using Bayesian networks, **Computer-Aided Civil and Infrastructure Engineering, Vol. 23, No. 3**, 189-207.

Castillo, E., Menéndez, J.M., Jiménez, P., (2008d) Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations. **Transportation Research-B, Vol. 42, No. 5**, 455-481.

Cremer, M. and Keller, H. (1987) A new class of dynamic methods for the identification of Origin-Destination flows, **Transportation Research-B, Vol. 21, No. 2**, 117-132.

Chang, G.L. and Wu, J. (1994) Recursive estimation of time-varying origin-destination flows from traffic counts in freeway corridors, **Transportation Research-B, Vol. 28, No.2**,

141-160.
Cho, H. J., (1991) Generalized inverse approach to sensitivity analysis of equilibrium network flow, **Transportation Planning Journal, Vol. 20, No. 1**, 1-44.
Cho, H. J., (1992) Comparison of minimum distance and generalized inverse approach for the sensitivity analysis of equilibrium network flow, **Transportation Planning Journal, Vol. 21, No. 1**, 23-34.
Davis, G.A. (1993) Estimating freeway demand patterns under impact of uncertainty on ramp controls, **ASCE Journal of Transportation Engineering, Vol. 119, No. 4**, 489-503.
Dixon, M.P. and Rilett, L.R. (2005) Population Origin-Destination estimation using automatic vehicle identification and volume data, **Journal of Transportation Engineering, Vol. 131, No. 2**, 75-82.
Eisenman, S.M. and List, G. F. (2004) Using probe data to estimate OD matrices, **Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference**, 291-296.
Friedberg, S. H., Insel A. J., and Spence, L. E., (2003) **Linear Algebra**, 4th Edition, New Jersey: Pearson Education.
Hellinga, B. and Van Aerde, M. (1994) A statistical analysis of the reliability of using RGS vehicle probes as estimators of dynamic O-D departure rates, **IVHS Journal, Vol. 2, No.1**, 21-44.
Horiguchi, R., Kuwanhara, M. and Akahane,H. (2001) The optimal arrangement of infrared beacons on a road network to collect vehicle trajectories-pattern analysis using schema theory, **Journal of the Eastern Asia Society for Transportation Studies, Vol. 4**, 219-228.
Hu, S.R., Peeta, S. and Chu, C.H. (2009a) Determination of a subset of links for sensor installation to infer network traffic flow information, **Proceedings of the 88th Transportation Research Board Annual Meeting**, Washington, D.C., U.S.A.
Hu, S.R., Peeta, S. and Chu, C.H. (2009b) Identification of vehicle sensor locations for link-based network traffic applications, **Transportation Research-B, Vol. 43, No. 8-9**, 873-894.
Hu, S.R. and Peeta, S. (2008) Inferring link flows using partial flow information: a strategic vehicle sensor deployment approach, **The Institute For Operations Research and The Management Sciences (INFORMS) Annual Meeting**, Washington, D.C., U.S.A.
Hu, S.R., Wang, C.Y. and Shih, C.J. (2008) Determination of network Origin-Destination demands using partial link traffic count information, **The 10th International Conference on Applications of Advanced Technologies in Transportation**, Athens, Greece.
Hu, S.R. and Wang, C.M. (2008) Vehicle detector deployment strategies for the estimation of network Origin-Destination demands using link traffic counts, **IEEE Transactions on Intelligent Transportation Systems**, **Vol. 9, No. 2**, 288-300.
Lam, W. H. K. and Lo, H. P. (1991) Accuracy of O-D estimates from traffic counts, **Traffic Engineering and Control, Vol. 31**, 358-367.
Leblanc, L. J. and Farhangian, K. (1982) Selection of a trip table which reproduces observed link flows, **Transportation Research-B, Vol. 16B, No. 2**, 83-88.
Lewis, C. D., (1982) **Industrial and business forecasting methods: a practical guide to exponential smoothing and curve fitting**, Butterworth Scientific, London.
Maher, M.J. (1983) Inferences on trip matrices from observations on link volumes: A Bayesian statistical approach, **Transportation Research-B, Vol. 17, No. 6**, 435-447.
Nihan, N. L. and Davis, G. A. (1987) Recursive estimation of Origin-Destination matrices from input/output counts, **Transportation Research-B, Vol. 21, No. 2**, 149-163.
Nihan, N. L. and Hamed, M. M. (1992) Fixed-point approach to estimating freeway Origin-Destination matrices and the effect of erroneous data on estimate precision, **Transportation Research Record 1357**, 18-28.

Nguyen, S. (1974) An algorithm for the traffic assignment problem, **Transportation Science, Vol. 8, No. 3**, 203-216.

Nguyen, S. (1977) **Estimating an OD matrix from network data: a network equilibrium approach**, University of Montreal Publication, No. 60.

Nguyen, S. (1984) **Estimation Origin-Destination matrices from observed flows, Transportation Planning Models**, Editor: M. Florian, Elsevier Science Publishers B.V., Netherlands, 363-380.

Penny, J. E. T. and Lindfield, G. R. (2000) **Numerical Methods Using Matlab**, 2th Edition, New Jersey: Prentice Hall.

Sherali H. D., Narayanan, A. and Sivanandan, R. (2003) Estimation of Origin-Destination trip-tables based on a partial set of traffic link volumes, **Transportation Research-B, Vol. 37, No. 9**, 815-836.

Spiess, H. (1987) A maximum likelihood model for estimating Origin-Destination matrices, **Transportation Research-B, Vol. 21, No. 5**, 395-412.

Suzuki, H., Nakatsuji, T., Tanaboriboon, Y., Takahashi, K., (2001) Dynamic estimation of origin-destination travel time and flow on a long freeway corridor, **Transportation Research Record 1739**, 67-75.

Van Zuylen, H. J. and Willumsen L. G.. (1980) The most likely trip matrix estimated from traffic counts, **Transportation Research-B, Vol. 14, No. 3**, 281-293.

Wu, J. and Chang, G. L. (1996) Estimation of time-varying origin-destination distributions with dynamic screenline flows, **Transportation Research-B, Vol. 30, No. 4**, 277-290.

Yang, H., Sasaki, T., Iida, Y. and Asakuba, Y. (1992) Estimation of Origin-Destination matrices from link traffic counts on congested networks, **Transportation Research-B, Vol. 26B, No. 6**, 417-434.

Yang, H. and Zhou, J. (1998) Optimal traffic counting locations for Origin-Destination matrix estimation, **Transportation Research-B, Vol. 32, No. 2**, 109-126.

Zhou, X. and Mahmassani, H. (2006) Dynamic origin-destination demand estimation using automatic vehicle identification data, **IEEE Transactions on Intelligent Transportation Systems, Vol. 7 No. 1**, 105-114.