Transportation Research Part B 88 (2016) 1–20

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

# Robust network sensor location for complete link flow observability under uncertainty

Xiangdong Xu[a], Hong K. Lo[b,*], Anthony Chen[a,c], Enrique Castillo[d]

[a] Key Laboratory of Road and Traffic Engineering, Tongji University, Shanghai 201804, China
[b] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong
[c] Department of Civil and Environmental Engineering, Utah State University, Logan, UT 84322-4110, USA
[d] Department of Applied Mathematics and Computational Sciences, University of Cantabria, Santander 39005, Spain

ARTICLE INFO

*Article history:*
Received 30 January 2014
Revised 14 March 2016
Accepted 15 March 2016
Available online 2 April 2016

*Keywords:*
Sensor location
Complete link flow observability
Measurement error
Robust optimization
Redundant optimization

ABSTRACT

The link observability problem is to identify the minimum set of links to be installed with sensors that allow the full determination of flows on all the unobserved links. Inevitably, the observed link flows are subject to measurement errors, which will accumulate and propagate in the inference of the unobserved link flows, leading to uncertainty in the inference process. In this paper, we develop a robust network sensor location model for complete link flow observability, while considering the propagation of measurement errors in the link flow inference. Our model development relies on two observations: (1) multiple sensor location schemes exist for the complete inference of the unobserved link flows, and different schemes can have different accumulated variances of the inferred flows as propagated from the measurement errors. (2) Fewer unobserved links involved in the nodal flow conservation equations will have a lower chance of accumulating measurement errors, and hence a lower uncertainty in the inferred link flows. These observations motivate a new way to formulate the sensor location problem. Mathematically, we formulate the problem as min–max and min–sum binary integer linear programs. The objective function minimizes the largest or cumulative number of unobserved links connected to each node, which reduces the chance of incurring higher variances in the inference process. Computationally, the resultant binary integer linear program permits the use of a number of commercial software packages for its globally optimal solution. Furthermore, considering the non-uniqueness of the minimum set of observed links for complete link flow observability, the optimization programs also consider a secondary criterion for selecting the sensor location scheme with the minimum accumulated uncertainty of the complete link flow inference.



## 1. Introduction

### 1.1. Research subject and motivation

The importance of link flow information for traffic management and the financial burden of installing sensors bring forth the need to effectively estimate all link flows from limited measurements in a network. The link flow observability problem

* Corresponding author. Tel.: +852 2358 8742.
*E-mail address:* cehklo@ust.hk (H.K. Lo).

http://dx.doi.org/10.1016/j.trb.2016.03.006
0191-2615/

is to identify the minimum number of sensors and their optimal locations such that the measurements collected allow the full determination of flows on all the unobserved links (Castillo et al., 2013b, 2015). The *minimum number* of sensors guarantees the lowest installation and operation cost, while the *optimal locations* ensure full observability, i.e., all unobserved link flows can be inferred from the observed flows. The full determination of all link flows is useful for numerous transportation applications, such as origin-destination (OD) trip table estimation, path flow estimation, network performance assessment, evacuation planning, and pavement management (see, e.g., Hu et al., 2009; Ng, 2012, 2013 and the references therein).

In reality, the observed link flows are inevitably subject to measurement error/inaccuracy. For example, Caltrans performance measurement system (PeMS) is a statewide repository for traffic data gathered by thousands of automatic sensors. Bickel et al. (2007) discussed detecting sensor malfunction, imputation of missing or bad data, estimation of velocity, and forecasting of travel times on freeway networks. They pointed out single-loop detectors are the main source of traffic data in California, but loop data are often missing or invalid due to communication error or hardware breakdown. A loop detector can fail in many ways even when it reports values. Payne et al. (1976) identified various types of detector errors, such as stuck sensors, hanging on or hanging off, chattering, cross-talk, pulse breakup, intermittent malfunction, etc. Even under normal conditions, loop detector measurements could be noisy, e.g., due to the confusion of multi-axle trucks. These measurement errors are not constrained to be localized. Instead, they can accumulate and propagate to affect the uncertainty of unobserved link flows, which subsequently can reduce the reliability of link flow inferences. Other than the source of measurement errors, the variability or uncertainty of the observed link flows also adds to the problem. In general, the measurements represent just a sample of the variable link flows. Using these sample measurements to infer the unobserved link flows adds to the uncertainty of the estimates. However, for both types of uncertainty sources, the location-specific degree of uncertainty for the observed link flows is *unavailable* before specifying the observed links and measuring their traffic counts. Our *study focus* is a new network/region, where there is no existing sensor and measurement information. Hence, an explicit uncertainty consideration in the optimization of traffic sensor locations for complete link flow observability is particularly crucial at the strategic planning stage of sensors installation, where observed link flows and their measurement errors are not yet available. We still can indirectly/implicitly consider the uncertainty propagation based on network topology when optimizing the traffic sensor locations without prior knowledge on the variance. This consideration will reduce the uncertainty in the unobserved link flows, while still ensuring the complete link flow observability in sensors location selection.

### *1.2. Classification of network sensor location problem*

Gentili and Mirchandani (2012) provided a detailed review on the network sensor location problem (NSLP) by categorizing it into two main types: sensor location *flow-observability* problem and sensor location *flow-estimation* problem. The former identifies the optimal placement of sensors that allows the unique determination of the unobserved flows based on the system of linear equations associated with the sensors. The latter identifies the optimal placement of sensors to best improve the quality of related estimates (e.g., OD demands, link/route flows) obtained by the system of linear equations associated with the sensors. Fig. 1 provides a selective summary of these two research problems with a focus on the uncertainty consideration, thus providing a comprehensive review is not the purpose of this study.

For the sensor location *flow-estimation* problem, majority of the literature focused on the link counting locations for OD estimation. Yang et al. (1991) first proposed the concept of maximal possible relative error to analyze the reliability of an OD matrix estimated from traffic counts. Yang and Zhou (1998) further derived four location rules: OD covering rule, maximal flow fraction rule, maximal flow-intercepting rule, and link independence rule. Ehlert et al. (2006) adopted the OD covering rule to develop the second best solution to locate additional counts with budget consideration. Yang et al. (2006) proposed the OD separation rule for the screenline-based traffic counting location problem. Chootinan et al. (2005b) considered the bi-objective counting location problem: the minimum number of counts to separate all OD pairs, and the maximal coverage for a given number of counts. Chen et al. (2007) adopted the OD separation rule and developed strategies for selecting additional counts for improving OD estimation. All of the above studies assumed that the traffic counts are error-free.

The *uncertainty* in sensor location *flow-estimation* problem has multiple sources, e.g., prior OD demand variability and measurement errors of traffic counts. Among others, Fei et al. (2007) and Fei and Mahmassani (2011) used the Kalman filtering method to identify a set of sensor locations that optimize the OD demand coverage and maximize the information gains through the observed data, while allowing for measurement errors and minimizing the uncertainties of the estimated OD demands in a time-varying context. Zhou and List (2010) optimized the locations of traffic counting stations and automatic vehicle identification (AVI) readers by maximizing the expected information gain for the OD estimation problem. Several error sources were considered, such as the uncertainty in historical demand information, measurement errors, and approximation errors of link flow proportions. The variability of the posterior OD estimate was measured through the trace of the covariance matrix. A linear measurement equation with an error term is used to relate the unknown OD demand to both point (e.g., link counts) and point-to-point (e.g., vehicle identification counts) measurements, where the distribution and covariance of measurement errors were explicitly considered. Simonelli et al. (2012) developed a network sensor location procedure by using a synthetic dispersion measure to quantify the variability of the posterior OD matrix estimate. A linear measurement equation under the error-free assumption was used to relate travel demand vector and link flow vector. Wang et al. (2012) maximized the variance reductions in posterior route flow estimates while considering the prior route flows and their reliabilities. Wang and Mirchandani (2013) optimized sensor locations using Bayesian inference to