Transportation Research Part B 105 (2017) 86–119

Contents lists available at ScienceDirect

Transportation Research Part B

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

Exact and approximate route set generation for resilient partial observability in sensor location problems

Marco Rinaldi*, Francesco Viti

University of Luxembourg, Research Unit of Engineering Science, Faculty of Science, Technology and Communication. 6, Rue R. Coudenhove-Kalergi, L-1359, Luxembourg

A R T I C L E   I N F O

Article history:
Received 7 April 2017
Revised 4 August 2017
Accepted 9 August 2017
Available online 21 September 2017

Keywords:
Network Sensor Location Problem
Partial link flow observability
Route set generation
Maximum clique problem
Genetic algorithm

A B S T R A C T

Sensor positioning is a fundamental problem in transportation networks, as the location of sensors strongly determines how traffic flows are observable and hence manageable.

This paper aims to develop a methodology to determine sensor locations on a network such that an optimal trade-off solution is found between the amount of sensors installed and the resilience of the sensor set.

In particular, we propose exact and heuristic solutions for identifying the optimal route sets such that no other route would include any additional information for finding optimal full and partial observability solutions. This is an important contribution to sensor location problems, as route-based link flow inference problems have non-unique solutions, strongly depending on the used link-route information.

The properties of the new methodology are analyzed and illustrated through different case studies, and the advantages of the algorithms are quantified both for full and for partial observability solutions. Due to the route sets found by our approach, we are able to find full observability solutions characterized by a small number of sensors, while yet being efficient also in terms of partial observability. We perform validation tests on both small and real-life sized network instances.

1. Introduction

Traffic information and management applications strongly rely on how traffic flows are monitored. Locating traffic sensors in a network is therefore considered a problem of paramount importance in transportation engineering, in particular within estimation problems (e.g. real time traffic state estimation, Ahmed et al., 2014; Zhu et al., 2014), OD flows estimation (Hadavi and Shafahi, 2016; Hu and Liou, 2014; Zhou and List, 2010), link flow inference (Castillo et al., 2008c; Hu et al., 2009; Liu et al., 2014; Xu et al., 2016), travel time estimation (Viti et al., 2008; Xing et al., 2013) and path flow reconstruction (Cerrone et al., 2015; Fu et al., 2016, 2017; Li and Ouyang, 2011).

Among others, network flow observability is a class of the so-called Network Sensor Location Problems (NSLP), in which the main goal is to determine the minimum set of observed link, route or OD flows that can be measured to provide information on the remaining non-observed link (or route, or OD) flows. Solutions to these problems are approached in literature by exploiting the fundamental relationships between the three sets of variables (link flows, route flows, OD flows), derived from conservation of vehicles principles. In this study we specifically focus on the *link flow inference* problem, i.e. the

* Corresponding author.
E-mail addresses: marco.rinaldi@uni.lu (M. Rinaldi), francesco.viti@uni.lu (F. Viti).

http://dx.doi.org/10.1016/j.trb.2017.08.007

problem of identifying a (smallest) set of independent links able to fully determine the flow on other links in the network (Castillo et al., 2015).

Existing approaches that compute link flow inference solutions are subdivided in methods exploiting node-link relations (Ng, 2012) or link-route relations (Castillo et al., 2008a; Hu et al., 2009). This latter category of approaches has been shown to potentially identify more efficient solutions. The work of Viti et al. (2014) showed that, for mid- and large-sized networks, node-based approaches tend to recommend a systematically higher number of sensors to install if compared to their route-based counterparts. A fundamental reason for this systematic difference was observed in the richer information offered by link-route relations if compared to link-node relations.

Despite their theoretical relevance, full observability solutions are however of little practicality in real-life networks, as the number of links to equip with sensors easily grows beyond economic feasibility with the size and complexity of the network itself. In Castillo et al. (2014) the authors reported that such solutions usually require around 60% of link flows to be observed, which is clearly not feasible in networks with hundreds or thousands of links. Given this argument, the concept of *partial* observability becomes very appealing. In partial observability problems, approaches must take into account different constraints (e.g. maximum number of sensors to be installed, location-specific restrictions, monetary budget limits, …), and therefore establish a trade-off between amount of measured/observed information and solution feasibility. The focus of the underlying problem shifts then from finding efficient measured link sets for complete link flow inference to, instead, maximizing the amount of *total* available information (different definitions of available information are provided later in Section 2 of this article). In Viti et al. (2014) we developed a partial observability metric to characterize and classify partial observability solutions, showing how this metric was capable to capture partial relationships and locate sensors in a very intuitive manner. While exploring the properties of this metric we quickly realized that algebraically equivalent full observability solutions actually exhibit strong behavioral differences when analyzed from a partial observability perspective.

More recently, in Rinaldi et al. (2015) we refined these results by empirically assessing the influence that route enumeration criteria have on the shape and structure of the full observability solution. We specifically concentrated on studying how the amount and quality of information in either full or partial information solutions depends on the number of routes and on the composition of the route set. The main finding of the paper was that the overall information content tends to increase as the route set is expanded, but the rate of growth reduces non linearly with the route set size, suggesting that there is an upper bound above which no new information is gained by including additional routes in the route set. Moreover, enumerating routes according to algebraic independence principles resulted in overall better-informative observability solutions with respect to standard enumeration techniques, a conclusion entirely in line with the findings of Castillo et al. (2014).

However, due to the combinatorial nature of the problem at hand, in our previous approaches we have been limited to showcasing distributions of information contents, while incapable of isolating the single best possible combination of route set and, thus, independent/dependent variable sets.

In this work we attempt to bridge this methodological gap, by indeed studying the nature of information in observability problems. Throughout the rest of the paper we will develop methodologies to exactly and approximately generate such a route set, and empirically verify whether indeed the resulting full observability solution is optimal in terms of partial observability information content.

The remainder of this work is structured as follows: in Section 2 we present a concise literature review related to Network Sensor Location Problems and, more specifically, to observability problems. In Section 3 we introduce an exact methodology for determining the maximum independent route set for any given network as well as a heuristic algorithm, devised to overcome some of the exact algorithm's computational limitations. Section 4 details two test cases, based on several small to mid-sized networks, aimed at verifying the key hypothesis introduced in Section 3. In Section 5 a meta-heuristic approach is developed, in order to extend the results and insights obtained for smaller networks towards real life instances. Section 6 presents validation results of the metaheuristic approach wrt. the exact results obtained in Section 4, and, finally, results on two mid-large sized networks. Finally, in Section 7, conclusions and remarks related to future research are discussed.

2. Literature review

2.1. Network Sensor Location Problems

The literature related to NSLP can be separated into approaches focusing on the algebraic and topologic properties of the network structure and connections (*observability problems*) and in those relating observed traffic states (usually, flows) with the ones derived using estimation techniques (*flow-estimation problems*) or including behavioral models and/or statistical models to predict future traffic states (*flow prediction problems*). For a more extensive overview of these problem types, their formalization and the relevant literature we refer to Gentili and Mirchandani (2012), Viti et al. (2014) and Castillo et al. (2015).

In observability problems, observed states from measurements can be used to infer this information to unobserved links states thanks to basic conservation of vehicles principles, i.e. when an unobserved variable can be related to only observed variables, it also becomes observable. In particular, in full observability solutions *all* unobserved flows are indirectly observable, as they all can be described as (linear) functions of the measured states. Flow-estimation problems, instead, seek to