# THE DETECTION LAYOUT PROBLEM

**Authors:**
**J.Barceló[1], F. Gilliéron, M.P. Linares[1], O. Serch[1], L.Montero[1]**

[1] Department of Statistics and Operations Research and CENIT
(Center for Innovation in Transport)
Technical University of Catalonia
Campus Nord, Building K2M, Office 306
Jordi Girona 29
08034 Barcelona
Spain

(jaume.barcelo, lidia.montero, mari.paz.linares,oriol.serch)@upc.edu
fanny.gillieron@gmail.com
Tel: +34 93 405 4659

Paper submitted for presentation and publication to
91st Transportation Research Board 2012 Annual Meeting
Washington, D.C.
November 2011

# WORDS: 6317

Paper revised from original submittal

## ABSTRACT

The primary data input used in principal traffic models comes from Origin-Destination (OD) trip matrices, which describe the patterns of traffic behavior across the network. In this way, OD matrices become a critical requirement in Advanced Traffic Management and/or Information Systems that are supported by Dynamic Traffic Assignment models. However, because OD matrices are not directly observable, the current practice consists of adjusting an initial or seed matrix from link flow counts which are provided by an existing layout of traffic counting stations. The adequacy of the detection layout strongly determines the quality of the adjusted OD. The usual approaches to the Detection Layout problem assume that detectors are located at network links. The first contribution of this paper proposes a modified set that formulates the link detection layout problem with side constraints. It also presents a new metaheuristic tabu search algorithm with high computational efficiency. The emerging Information and Communication Technologies, especially those based on the detection of the electronic signature of on-board devices (such as Bluetooth devices) allow the location of sensors at intersections. To explicitly take into account how these ICT sensors operate, this paper proposes a new formulation in terms of a node covering problem with side constraints that, for practical purposes, can be efficiently solved with standard professional solvers such as CPLEX.

**Keywords:** OD Estimation, Dynamic User Equilibrium, Traffic Detector Location, Link Covering, Node Covering, Integer Programming, Location Theory

## THE DETECTION LAYOUT PROBLEM

Traffic detectors are usually the means for measuring fundamental traffic variables (i.e. flows, speeds and occupancies), whose values determine the state of the traffic system. The values of these variables are the main inputs to models based on traffic flow theory for estimating the state of traffic flow at freeway sections, either to study and understand traffic behavior or to support efficient traffic management decisions. These values are also the input to traffic control systems which are used to calculate traffic control plans --be they fixed, actuated or adaptive-- in order to optimize the traffic network. Traffic detectors are usually located along the road in the best positions for achieving these goals. However, when dealing with traffic networks, namely complex urban networks, additional components are necessary for describing and understanding the traffic patterns which account for the behavior of traffic flows in the network. Origin-to-Destination trip matrices describe the number of trips between any origin-destination pair in a traffic network. Route choice models describe how trips select the available paths between origins and destinations and, as a consequence, the number of trips using a given path. In other words, they describe the path flows or the path flow proportions, depending on whether we refer to the number of trips using a path or to the fraction of trips using a path, with respect to the total number of trips between the corresponding origin and destination.

Traffic assignment models have the objective of assigning a trip matrix onto a network --in terms of a route choice mechanism-- in order to estimate the traffic flows in the network. Therefore, they all use Origin-Destination (OD) trip matrices as major data input for describing the patterns of traffic behavior across the network. All formulations of static traffic assignment models (Florian and Hearn [1], as well as dynamic, Ben-Akiva et al. [2]) assume that a reliable estimate of an OD is available. However, OD matrices are not yet directly observable, even less so in the case of the time-dependent OD matrices that are necessary for Dynamic Traffic Assignment models; consequently, it has been natural to resort to indirect estimation methods. These indirect estimation methods are the so-called matrix adjustment methods, whose main modeling hypothesis can be stated as follows: if traffic flows in the links of a network are the consequence of the assignment of an OD matrix onto a network, then, if we are capable of measuring link flows, the problem of estimating the OD matrix that generates such flows can be considered as the inverse of the assignment problem (Cascetta [3]). Since the earlier formulation of the problem by Van Zuylen and Willumsen [4], the matrix adjustment problem has been a relevant research and practical problem.

The current practices consist of using an initial OD estimate, the OD seed or OD target as input, and adjusting them from the available link counts provided by an existing layout of traffic counting stations and other additional information whenever it is available. Adjustments can be considered as indirect estimation methods, based either on discrete time optimization approaches (Codina and Barceló [5], Lundgren and Peterson [6]) or on adaptations of Kalman Filtering approaches (Ashok et al. [7], Antoniou et al. [8], Barceló et al. [9]). In summary, the modeling approaches change depending on whether the adjustment is based on static or time-dependent formulations of the underlying traffic assignment problem; but they all share two fundamental modeling hypotheses:

1. That a mapping scheme of OD flows-link flow counts is available
2. That, if $A$ is the set of links in the network, flow detectors are only located in a subset $\hat{A} \subset A$ of links in the network, from which link flow measurements $\hat{v}_a$ , $a \in \hat{A}$ are available

It soon becomes evident that, because they are designed and implemented with the primary purpose of providing the data required by traffic control applications, the current detection layouts in traffic networks are not appropriate for the reconstruction of OD matrices, because they do not take explicitly into account the OD pattern structure. The objective of identifying a detection layout that optimizes the coverage of origin-destination demand on the road network while minimizing the uncertainties of the estimated OD matrix still remains an important challenge for Advanced Traffic Management Systems, as well as for the transportation studies necessary for the design, feasibility and impact evaluation of such systems. The main research on this topic assumes that traffic detectors (for instance inductive loops) are usually located in the network links. This paper explores a modified set which covers the formulation of the classical link detection layout problem, including side constraints that model specific conditions for achieving the objectives, and it also develops a new metaheuristic algorithm whose computational efficiency is tested with various real networks. However, we believe that the possibilities raised by the emerging Information and Communication Technologies cannot be ignored. This situation creates new scenarios. For example, layouts based on the detection of the electronic signature of on-board devices, such as Bluetooth devices, allow the location of sensors at intersections. To explicitly take into account how these ICT sensors operate, this paper proposes a new formulation in terms of a node-covering problem with side constraints that, for practical purposes, can be efficiently solved with standard professional solvers such as CPLEX.

## FORMULATIONS OF THE LINK DETECTION LAYOUT PROBLEM

To understand the role of the detection layout in the estimation of the OD matrix, it is crucial to understand the relationships between link flows, path flows and the OD matrix. In the case of time independent approaches --that is, when the OD matrix is assumed to be constant for the time period under consideration-- models that adjust OD flows from link flow counts assume a mapping of OD flows to counts in which, if $N_I$ is the number of OD pairs in the network, $N_A$ is the number of arcs and $N_K$ is the number of paths between OD pairs, then the matrix $P[N_K, N_I]$ of path choice fractions is known or can be computed. Each entry $f_{ik}$ of this matrix is the fraction of trips of the i-th OD pair using path $k$. Given the set K of used paths, the link-path incidence matrix $\Delta[N_A,N_K]$ between links $a \in A$ of the network and paths $k \in N_K$ is defined as:

$$\delta_{ak} = \begin{cases} 1 \; if \; link \; a \; is \; on \; path \; k \\ 0 \; otherwise \end{cases}$$

The Assignment Matrix $A[N_A, N_I]$ is defined as $A=\Delta P$ and therefore the fraction of trips for the i-th OD pair using link $a$ is $p_{ia} = \sum_k f_{ki}\delta_{ak}$ . Thus, the flow on link $a$ is given by:

$$v_a = \sum_{i \in I} p_{ia} g_i \qquad (1)$$

Where $g_i$ is the total number of trips for the $i$-th OD pair. This mapping, between the OD matrix $g$ and the link flows $v_a$ , is the basic constraint in mathematical programming models for OD adjustment from link flow counts. Assuming that flow counting stations are located in a subset of links $\hat{A} \subset A$ and that $\hat{v}_a$ is the measured flow on link $a$, then, in accordance with (1), the following relationship holds:

$$\sum_{i \in I} p_{ia} g_i = \hat{v}_a \, , \forall a \in \hat{A} \qquad (2)$$

Depending on whether proportions $p_{ia}$ are considered constant or a function of the OD matrix $g$ being adjusted, different algorithmic approaches toward $p_{ia}$ can be proposed (see Lundgren and Peterson [6] for details).

When traffic demand is time–dependent, the static model in (2) is modified to account for time dependencies. In this case, most of the models formulate the adjustment problem in terms of autoregressive approaches (Ashok et al. [7], Antoniou et al. [8], Barcelo et al. [9]) and the fundamental relationships between traffic counts and OD flows can then be stated as follows (Ashok el al. [7]):

$$v_{ah} = \sum_{r=h-r'}^{h} \sum_{i=1}^{n_{OD}} p_{ah}^{ir}\, g_{ih} + \varepsilon_{ah} \tag{3}$$

Where $v_{ah}$ are the measured traffic counts at a given detector on link *a* during time interval $h$, $p_{ah}^{ir}$ is the fraction of the *i*-th OD flow that departed from its origin during interval *r* and crossed the detector at link *a* during interval $h$. $g_{ih}$ is the number of vehicles between the i-th OD pair that left their origin in interval $h$, $\varepsilon_{ah}$ is the measurement error and $r'$ is the order of the autoregressive model. $p_{ah}^{ir}$ generalizes the concept of mapping OD flows and link counts.

The quality of the adjusted matrix strongly depends on: the quality of the seed matrix in reducing the indetermination of the problem, the quality of the mapping schemes and the quality of the detection layout. Bierlaire [10] analyzes these dependencies for a given layout. Although it was early evidenced that the role of constraints (2) and (3) depend not only on the number of detectors but also on their location in the network, the problem has not received substantial attention until recently. The advent of Advanced Traffic Information and Active Traffic Management Systems (ATIS and ATMS respectively) has fostered the use of Dynamic Traffic Assignment models whose main input are OD matrices and, therefore, the need for reliable OD matrices has brought the Detection Layout Problem into the forefront.

From the point of view of the estimation of time-dependent OD matrices, the optimal sensor location is a problem strongly related to the observability of the network. Castillo *et al.* [11] define the observability problem as the problem of identifying whether a set of available measurements is sufficient for estimating the state of a system. In the case of a state representation approach, this means that --for any possible sequence of state and control vectors-- the current system state (and therefore, the system behavior) can be determined using only the system measurements. Castillo *et al.* [11] reformulate the problem for traffic networks in terms of "determining if a set of OD pair and link flows is sufficient to estimate the state of the network".

Yang and Zhou [12] proposed a formulation in 1998 of the problem in terms of a set covering model with additional constraints, which is still considered the main reference. They formulate the model on the basis of four modeling hypotheses or basic rules that must be satisfied by any optimal covering:

1. O/D Covering rule: The counting stations on the road network should be located so that a certain proportion of trips between any O/D pair will be observed.
2. Maximal Flow Fraction rule: For a particular O/D pair, the counting stations on the road network should be located at the links so that the flow fraction between this O/D pair over flows on these links is as large as possible.
3. Maximal Flow Intercepting rule: Under a certain number of links to be observed, the chosen links should intercept as many flows as possible.

4. Link Independence rule: The counting stations should be located on a network so that the resultant traffic counts on all chosen links are not linearly dependent.

As a consequence of Rule 4, a primary formulation of the problem follows. Let:

$$x_a = \begin{cases} 1 & if\ a\ counting\ station\ is\ located\ on\ link\ a \\ 0 & otherwise \end{cases}$$
$$\delta_{ai} = \begin{cases} 1 & if\ some\ trips\ between\ OD\ pair\ i\ pass\ over\ link\ a \\ 0 & otherwise \end{cases}$$

Then, the following combinatorial optimization model solves the formulation of the location problem that satisfies the Link Independence Rule (Rule 4):

$$\begin{aligned} & MIN\ x_0 = \sum_{a \in A} x_a \\ & s.t. \\ & \quad \sum_{a \in A} \delta_{ai} x_a \geq 1, \forall i \in I \\ & \quad x_a \in \{0,1\}\ \ \forall a \in A \end{aligned} \tag{4}$$

A more complete formulation can be derived including Rule 3. Let's define:

$$y_k = \begin{cases} 1 & if\ there\ is\ at\ least\ one\ counting\ station\ along\ path\ k \\ 0 & otherwise \end{cases}$$

$$\hat{l} = maximum\ number\ of\ counting\ stations\ to\ be\ located$$

$$h_k = flow\ on\ path\ k \in K$$

The detection layout problem can then be reformulated in terms of an enhanced model that accounts for Rules 3 (Maximal Flow Intercepting Rule) and 4 (Link Independence Rule):

$$\begin{aligned} & MAX \sum_{k \in K} h_k\, y_k \\ & s.t. \\ & \quad \sum_{a \in A} x_a = \hat{l} \\ & \quad \sum_{a \in K} x_a \geq y_k, \forall k \in K \\ & \quad \sum_{a \in A} \delta_{ai}\, x_a \geq 1, \forall i \in I \\ & \quad x_a \in \{0,1\}, \forall a \in A;\ y_k \in \{0,1\}, \forall k \in K \end{aligned} \tag{5}$$

In practice, budgetary limitations can impose a bounding constraint on the maximum number $\hat{l}$ of detection stations that can be located on a network. In this formulation, Rule 3 is the objective of the optimization whereas Rule 4 is set as a constraint. The objective function computes the net captured flow. The problem can be infeasible if $\hat{l} < \hat{l}_0$, where $\hat{l}_0$ is the minimum number of detectors that make it possible to satisfy the OD covering constraints.

Variants of these formulations can be found in Yang's paper [12]. An analysis of the advantages and disadvantages of these formulations in terms of the quality of the adjusted OD matrix can be found in the paper by Larsson et al. [13].

A common drawback of these formulations is that they require a complete enumeration of all paths between all OD pairs in the road network, which leads to a problem in which the size grows exponentially with the size of the network. A more efficient formulation could be obtained by relaxing the requirement of explicitly enumerating the set of paths and using instead the subset of the most likely used paths. Relaxed formulations of the problem have recently received the attention of researchers (Ehlert et al. [14] and Xiang Fei et al. [15]).

## A MODIFIED FORMULATION OF THE LINK DETECTION LAYOUT PROBLEM ACCOUNTING FOR TIME DEPENDENCIES

On the basis of the previous discussion, we restrict the path set to the most likely used paths in order to reduce the size of the problem and achieve a higher computational efficiency while not significantly degrading the quality of the solution. Given that the objective of the optimization model is maximum flow interception, and that the most likely used paths are the ones expected to accommodate the greatest number of trips, these paths should be covered even if other paths are used by a minority of travelers

The approach taken in this paper can be considered as an extension of a previous exploration of the detection layout problem made by Hoogland [16], who explored two alternative formulations: one based on a proposal by Bianco et al. [17] which is rooted in the topological analysis of the network, and another one in which the identification of the most likely used paths for reducing problem size was based on a static equilibrium assignment performed by the Emme commercial package. In the solution produced from the latter, the paths were identified according to a post-processing procedure based on Larsson et al. [18]. A drawback of this approach was that --being based on a static equilibrium assignment-- it did not account for the time dependencies of the used paths. This motivated our research to try and capture this time-dependent nature by identifying the most likely used paths on the basis of a heuristic, dynamic user equilibrium traffic assignment with Dynameq [17]. Fei et al. [20] also exploit the advantages of formulating the problem in terms of the most likely used paths as determined by the dynamic user equilibrium achieved by DYNASMART, but their formulation is based on the simplified model discussed in (4). An additional feature of a formulation based on a dynamic traffic assignment is that, due to explicitly accounting for time dependencies, the identified paths and the corresponding link proportions can be time sliced, leading to the following enriched time-dependent formulation of the detection layout problem:

$$
\begin{aligned}
& MAX \sum_{\tau \in T} \sum_{k \in K} h_k^{\tau} y_k \\
& \sum_{a \in A} x_a = \hat{l} \\
& \sum_{a \in k} x_a \geq y_k,\ \forall k \in K \\
& \sum_{a \in A} \delta_{ai}^{\tau} x_a \geq 1,\ \forall i \in I, \forall \tau \in T \\
& x_a \in \{0,1\}, \forall a \in A,\ y_k \in \{0,1\}, \forall k \in K
\end{aligned}
\tag{6}
$$

Where $h_k^\tau$ is the flow on path k at time interval $\tau$, $\delta_{ai}^\tau$ indicates if link a is part of one of the most likely paths between OD pair $i$ at time interval $\tau$. With this restriction, we have a maximum number of paths that is equal to $N_I$x|T|xM, where M is the maximum number of paths that are considered for each OD pair for each time interval (usually 3 or 4) and |T| is the number of time intervals. This number can be great, but does not grow exponentially with the number of links in the network, and therefore is acceptable even for large networks. The resulting set covering problem with side constraints has a richer and improved structure compared to the previous one studied in [16]. From a computational point of view, the main difference lies in the fact that the resulting formulation can be efficiently pre-processed to reduce its size by using the pre-processing techniques that are proper for combinatorial optimization (Savelsbergh [20]). This allows the identification of dominated columns and rows, clique inequalities and other pertinent characteristics in the combinatorial structure of the set of constraints (6). The equivalent pre-processed problem is then solved by an ad hoc combination of tabu search and diversification heuristics (Hertz et al. [21]) in a modified problem formulation, which subsequently relaxes the OD covering rule in a Lagrangian fashion.

The tabu search is a local search method developed by Glover [22]. This algorithm tries to improve the current solution by choosing the best in the neighborhod, but it avoids some specific movements in order to increase the chances of leaving a local optimum. The algorithm keeps a list of forbidden movements --called a tabu list-- at each iteration; the next solution is then given by the best neighbor solution that does not need to reach a tabu movement. When the movement iscompleted, the inverse movement is added to the tabu list and will stay inside it for a fixed number of iterations. This will force the algorithm to search further, even if it is necessary to go through worse solutions. To avoid being too restrictive, a tabu movement will nevertheless be authorized if it reaches a solution that is better than any other found in the entire algorithm.

An algorithm capable of finding a good solution for any number of detectors is desirable, but it is not possible to know if this number is great enough to satisfy the OD covering rule or not. Even if it is possible to find a feasible solution, it could be less desirable to satisfy the OD covering rule while intercepting a small fraction of the total flow instead of not covering a few OD pairs while intercepting almost all the flow in the network. Following these considerations, the choice was made to no longer consider the OD covering rule as a constraint, but to introduce it instead into the objective function, which permits not only dealing with the fact that these constraints may not be feasible, but it also gives a direct means for evaluating the quality of a solution. Indeed, it is necessary to evaluate whether a solution that covers more OD pairs but less flow is better or worse than another solution. This has to be done by weighting what is more important: to satisfy Rule 1 (OD covering rule) by weighting the coefficient $\rho_2$, or to follow Rule 3 (maximal flow intercepting rule) by weighting coefficient $\rho_1$. The problem was therefore reformulated as follows:

$$\begin{aligned} & MAX \left\{ \rho_1 \frac{Intercepted\ Flow}{Total\ Flow} + \rho_2 \frac{\#\, OD\ Constraints\ Satisfied}{\#\, OD\ Constraints} \right\} \\ & s.t. \quad \sum_{a\in A} x_a = \hat{l} \\ & \qquad \sum_{a\in k} x_a \geq y_k, \forall k \in K \\ & \qquad x_a \in \{0,1\}, \forall a \in A, y_k \in \{0,1\}, \forall k \in K \end{aligned} \tag{7}$$

Where the intercepted flow is given by $\sum_{\tau \in T} \sum_{k \in K} h_k^\tau y_k$

The OD covering rule constraint $\sum_{a \in A} \delta_{ai}^\tau x_a \geq 1, \forall i \in I, \forall \tau \epsilon T$ is replaced by another one, using the path variables instead of the link variables:

$$\sum_{k \in K} \delta_{ki}^\tau y_k \geq 1, \forall i \in I, \forall \tau \in T$$

Where $\delta_{ki}^\tau \begin{cases} = 1 \; if \; path \; k \; connects \; the \; i - th \; OD \; pair \\ = 0 \; Otherwise \end{cases}$. This constraint still represents the OD covering rule, but now the left side member contains only a few variables, as the maximum number of paths for each OD pair is fixed.

**Greedy algorithm for initialisation**

The initial solution is found by the following greedy algorithm:

```
1. Initialisation
   Compute total flows on each link.
2. While some detectors aren't placed and some links have non-zero flow, do
    (a) Find link with greatest flow, add a detector to it.
    (b) For each path going through this link:
         set flow to zero.
         change link flows according to it.
3. Return to detector’s location
```

This greedy algorithm gives a solution that covers a great fraction of total flow, and therefore is a good start for the tabu search; but it is not optimal, neither for the number of OD constraints nor the flow.

**Next movement's choice**

A movement from one solution to another is defined as the move of a detector to a new location, which is on the one hand relatively simple to try and find, and on the other hand allows us to reach every possible solution. As a criterion for measuring a movement's quality, the improvement of the objective function (7) is considered. In order to find the best movement, detectors must be treated one by one, first by removing it and then by trying to insert a new detector located at each empty link.

```
1. Given:
   Set of links containing detectors
   Updated data for current solution with flows and constraints
   Evaluation function
2. For each detector in the set, do:
   (a) Remove detector and update data
   (b) If new flow on link where the detector was located is 0:
        Choose best available link to put detector.
        STOP (the detector was useless).
   (c) Else
        For each link without detector
        Evaluate movement with evaluation function.
```

```
            Keep if best.
        3. Do best movement and update data.
```

**Tabu List Update**

```
        1. Given:
           TabuList containing elements of the type (start; end; iteration), where
            start is the link where the detector was, end is the link where detector is
            placed, iteration is a step counter
           TabuListSize which represent the oldest movement that is considered in
            tabu list
            New movement m = (start; end; iteration)
        2. Add inverse movement (end; start; iteration) to TabuList
        3. For each movement m' in TabuList, do
           (a) If m'.iteration = m.iteration _ TabuListSize
                Remove m' from TabuList
           (b) Else if m'.start = m.end
                Add movement (m.start;m'.end;m'.iteration) to TabuList
        4. Return TabuList
```

With this definition of movement, two neighbor solutions are very close. In order to explore a large part of the solution space, we have to add a diversification phase to our algorithm, because the tabu list can't be big enough to ensure diversification, as it has to remain reasonable in order to intensively explore the neighborhood. When the current solution isn't improved during the last iterations of k, we temporally change the objective function to $MIN\ \sum_{a\in A} n(a)x_a$, where $n(a)$ is a frequency counter for placing a detector on link $a$, in order to reach an unexplored area. The tabu search then starts again.

## COMPUTATIONAL RESULTS FOR THE LINK LAYOUT MODEL

A set of computational experiments has been conducted with various networks. Due to the limitation of space, we illustrate the results by presenting here only those which correspond to the Eixample district in Barcelona, with the infrastructure corresponding to May 2005. This network, depicted in Figure 1, has 1570 sections, 692 intersections and 210 centroids. The Origin-Destination matrix we started with corresponds to the traffic between 8 am and 9 am. The OD matrix for this network was also split into four different matrices corresponding to four different time slices, using the following percentages: 15%, 30%, 40% and 15%. There were initially 1570 links that were reduced to 1289 by preprocessing; 531 clique inequalities were found. The equilibrium gave 8242 paths, reduced to 3107, for 1358 OD pairs during 4 time periods, which means a total number of 4944 initial OD constraints. This number falls down to 2609 when not considering the repeated constraints, and to 2045 when suppressing the redundant ones. The total flow assigned to the Barcelona network was 63973 trips, and the minimum number of detectors to cover the 1236 OD pairs is 115. Table 1 summarizes the numerical results and the red stars in Figure 1 identify the proposed detector locations. Our heuristic intercepts a large fraction of total flow (63973 vehicles), but also covers almost each OD pair in the network, which is important for limiting the OD matrix estimation error.

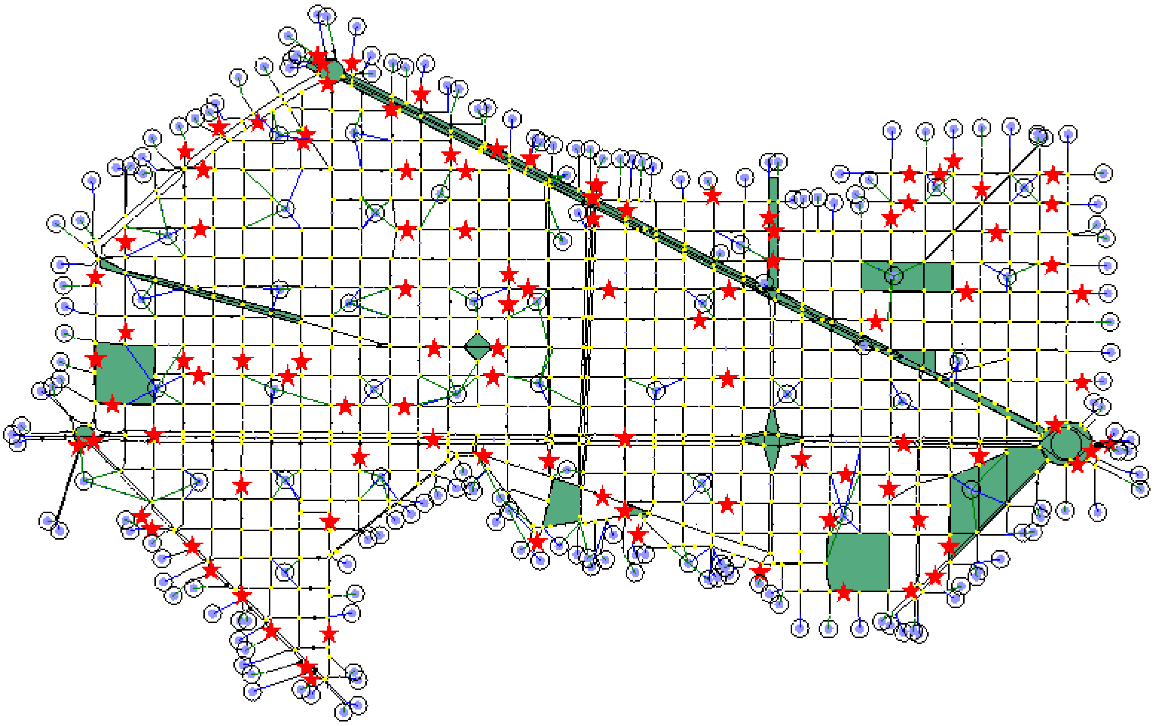


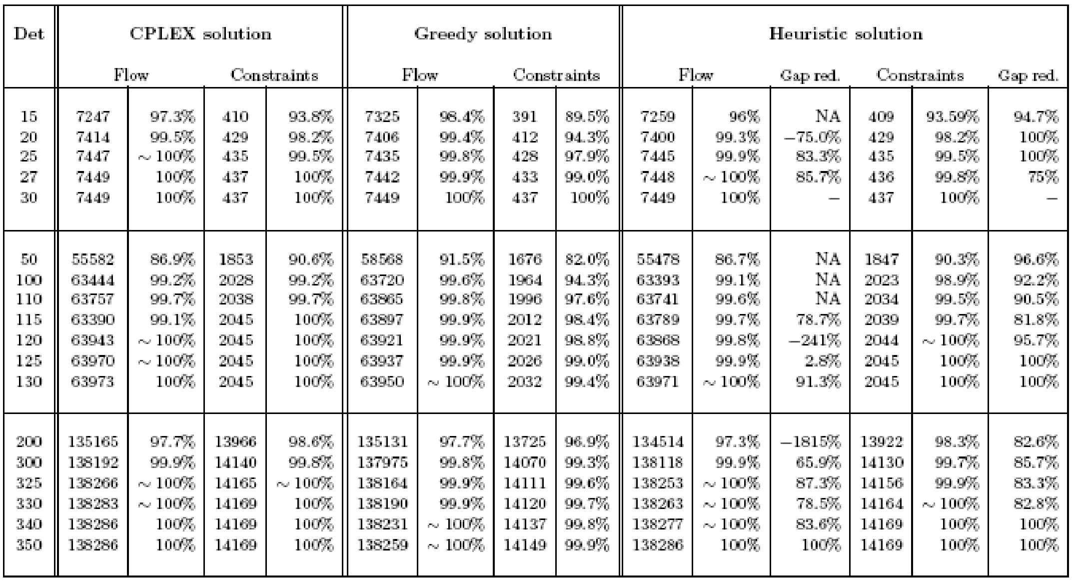
**Figure 1 - Detection layout for Barcelona's model**

| Detectors | Theoretical solution : intercepted flow | Constraints satisfied | Heuristic solution : flow intercepted | Constraints satisfied |
|---|---|---|---|---|
| 100 | 63890 (99.9%) | 1975/2045 | ~63490 (99.2%) | ~2022/2045 |
| **115** | **63390 (99.1%)** | **2045/2045** | **~63780 (99.7%)** | **~2039/2045** |
| 120 | 63943 (~100%) | 2045/2045 | ~63850 (99.8%) | ~2043/2045 |
| 130 | 63973 (100%) | 2045/2045 | ~63970 (~100%) | 2045/2045 |

**Table 1 - Computational results for Barcelona's network**

Table 2 reports on the quality of the computational results with a comparison of the problem's exact solutions that were obtained by CPLEX. The heuristic was modified to account for additional practical considerations, like limiting a priori the number of detectors to locate, a constraint that could be imposed in practice by budgetary conditions. Another practical limitation would be bounding the percentage of the total number of trips intercepted by the detectors, a constraint that could be imposed in certain conditions as a measure of the quality of the solution. Two other networks were used:

- The network of the City of Preston, in Lancashire UK, containing 417 links, 166 nodes (intersections) and 34 centroids representing origins and/or destinations.
- And the highway network of the Hessen land in Germany, with 4282 links (road sections), 495 nodes (intersections) and 245 centroids (origins/destinations).



|  | Det | CPLEX solution |  |  |  | Greedy solution |  |  |  | Heuristic solution |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  |  | Flow |  | Constraints |  | Flow |  | Constraints |  | Flow |  | Gap red. | Constraints |  | Gap red. |
| Preston | 15 | 7247 | 97.3% | 410 | 93.8% | 7325 | 98.4% | 391 | 89.5% | 7259 | 96% | NA | 409 | 93.59% | 94.7% |
|  | 20 | 7414 | 99.5% | 429 | 98.2% | 7406 | 99.4% | 412 | 94.3% | 7400 | 99.3% | −75.0% | 429 | 98.2% | 100% |
|  | 25 | 7447 | ∼ 100% | 435 | 99.5% | 7435 | 99.8% | 428 | 97.9% | 7445 | 99.9% | 83.3% | 435 | 99.5% | 100% |
|  | 27 | 7449 | 100% | 437 | 100% | 7442 | 99.9% | 433 | 99.0% | 7448 | ∼ 100% | 85.7% | 436 | 99.8% | 75% |
|  | 30 | 7449 | 100% | 437 | 100% | 7449 | 100% | 437 | 100% | 7449 | 100% | – | 437 | 100% | – |
| Barcelona | 50 | 55582 | 86.9% | 1853 | 90.6% | 58568 | 91.5% | 1676 | 82.0% | 55478 | 86.7% | NA | 1847 | 90.3% | 96.6% |
|  | 100 | 63444 | 99.2% | 2028 | 99.2% | 63720 | 99.6% | 1964 | 94.3% | 63393 | 99.1% | NA | 2023 | 98.9% | 92.2% |
|  | 110 | 63757 | 99.7% | 2038 | 99.7% | 63865 | 99.8% | 1996 | 97.6% | 63741 | 99.6% | NA | 2034 | 99.5% | 90.5% |
|  | 115 | 63390 | 99.1% | 2045 | 100% | 63897 | 99.9% | 2012 | 98.4% | 63789 | 99.7% | 78.7% | 2039 | 99.7% | 81.8% |
|  | 120 | 63943 | ∼ 100% | 2045 | 100% | 63921 | 99.9% | 2021 | 98.8% | 63868 | 99.8% | −241% | 2044 | ∼ 100% | 95.7% |
|  | 125 | 63970 | ∼ 100% | 2045 | 100% | 63937 | 99.9% | 2026 | 99.0% | 63938 | 99.9% | 2.8% | 2045 | 100% | 100% |
|  | 130 | 63973 | 100% | 2045 | 100% | 63950 | ∼ 100% | 2032 | 99.4% | 63971 | ∼ 100% | 91.3% | 2045 | 100% | 100% |
| Hessen | 200 | 135165 | 97.7% | 13966 | 98.6% | 135131 | 97.7% | 13725 | 96.9% | 134514 | 97.3% | −1815% | 13922 | 98.3% | 82.6% |
|  | 300 | 138192 | 99.9% | 14140 | 99.8% | 137975 | 99.8% | 14070 | 99.3% | 138118 | 99.9% | 65.9% | 14130 | 99.7% | 85.7% |
|  | 325 | 138266 | ∼ 100% | 14165 | ∼ 100% | 138164 | 99.9% | 14111 | 99.6% | 138253 | ∼ 100% | 87.3% | 14156 | 99.9% | 83.3% |
|  | 330 | 138283 | ∼ 100% | 14169 | 100% | 138190 | 99.9% | 14120 | 99.7% | 138263 | ∼ 100% | 78.5% | 14164 | ∼ 100% | 82.8% |
|  | 340 | 138286 | 100% | 14169 | 100% | 138231 | ∼ 100% | 14137 | 99.8% | 138277 | ∼ 100% | 83.6% | 14169 | 100% | 100% |
|  | 350 | 138286 | 100% | 14169 | 100% | 138259 | ∼ 100% | 14149 | 99.9% | 138286 | 100% | 100% | 14169 | 100% | 100% |

**Table 2 - Comparative results of the exact and heuristics solutions with additional constraints**

Table 2 compares the exact solution with CPLEX, the greedy heuristic and the tabu heuristic in terms of the number of detectors, the % of the total flow intercepted and the % of OD constraints satisfied in model (7).

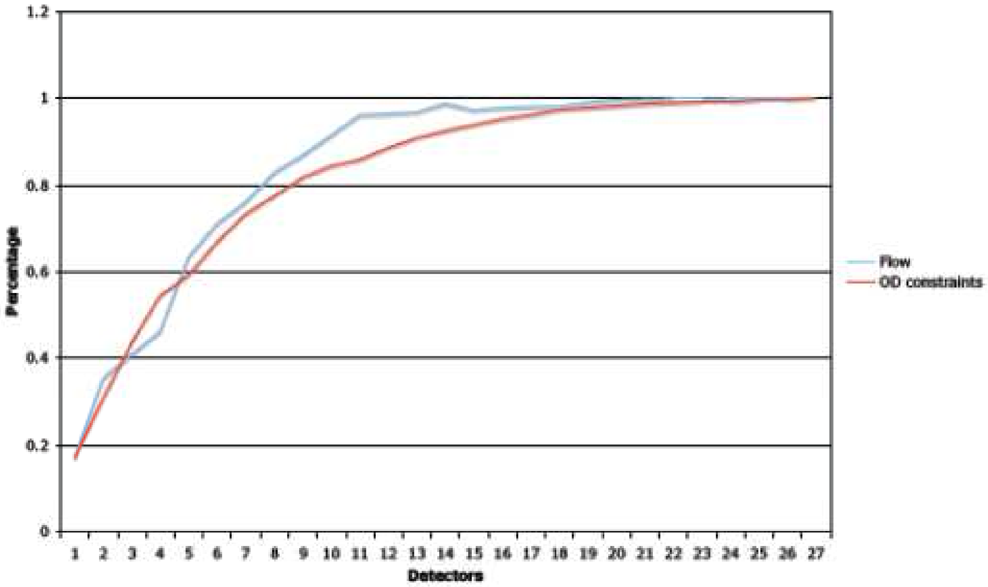


**Figure 2 - Covered flow and OD constraints in Preston network**

Figure 2 graphically illustrates the quality of the solution in terms of % of intercepted flows and % of satisfied OD constraints as a function of the number of detectors. It can be observed that both the flow and the constraints covered are improved very quickly with a small number of detectors. They then need a large number of additional counts to reach a total cover. The same phenomenon is observable in the Barcelona and Hessen networks.

## THE INTERSECTION DETECTION LAYOUT PROBLEM

Most of the new applications based on Information and Communications Technology (ICT) work in a different way from the traditional ones; they are able to capture the electronic or magnetic signature of specific on-board devices. One of the most typical of such sensors is that which is capable of capturing a Bluetooth equipped device on board a vehicle. The basic principles on how these sensors operate are the following: A vehicle equipped with a Bluetooth device traveling along the freeway is logged and time-stamped at time $t_1$ by the sensor at location 1. After traveling a certain distance it is logged and time-stamped again at time $t_2$ by the sensor at location 2 downstream. The difference in time stamps $\tau = t_2 - t_1$ measures the travel time of the vehicle equipped with that mobile device. Obviously the speed is also measured, assuming that the distance between both locations is known. Data captured by each sensor is sent for processing to a central server by wireless telecommunications.

However, when dealing with sensors that capture the electronic signature as described, and specifically when these are detectors of Bluetooth devices on board vehicles, the observability problem in terms of detector location must be formulated in different terms by taking into account that these detectors are more efficiently located at intersections and not at links, where they can capture a higher number of vehicles. Let's analyze the scheme in Figure 3, assuming that the Bluetooth sensor is located at the intersection in a location in such a way that its detection lobule intercepts all equipped vehicles crossing the node on paths (1), (2), (3) and (4). The candidate intersections would be those intercepting a higher number of equipped vehicles.

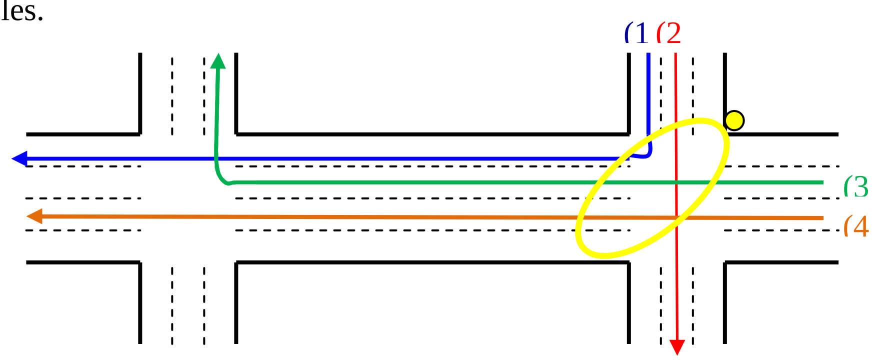


**Figure 3 - Flows intercepted from paths crossing a node**

Following the same four modeling hypotheses proposed in the formulation of the link detection layout problem, and as a consequence of Rule 3 (Maximal Flow Intercepting Rule), a primary formulation of the intersection layout problem is the following. Let:

$$x_n = \begin{cases} 1 & \mathit{if\ a\ sensor\ is\ located\ at\ intersection\ n} \\ 0 & \mathit{otherwise} \end{cases}$$
$$y_k = \begin{cases} 1 & \mathit{if\ there\ is\ at\ least\ one\ detector\ along\ path\ k} \\ 0 & \mathit{otherwise} \end{cases}$$
$$\delta_{nk} = \begin{cases} 1 & \mathit{if\ intersection\ n\ is\ into\ path\ k} \\ 0 & \mathit{otherwise} \end{cases}$$

$$\hat{l} = \mathit{maximum\ number\ of\ detectors\ to\ be\ located}$$
$$K = \mathit{set\ of\ all\ paths\ between\ all\ OD\ pairs}$$
$$K_i = \mathit{set\ of\ paths\ for\ the\ ith\ OD\ pair};\ K = \bigcup_{i\in I} K_i$$
$$h_k = \mathit{flow\ on\ path}\ k \in K$$
$$N = \mathit{set\ of\ intersections\ (nodes) of\ the\ network}$$
$$I = \mathit{set\ of\ all\ OD\ pairs\ in\ the\ network}$$

If path $k$ carries flow $h_k$ , then $\sum_{k\in K_n} h_k$ is the total flow captured by the detector at node n. $K_n$ is the set of paths crossing node n. Based on the experience with the previous link covering models, the problem is formulated in terms of the most likely used paths determined by the solution of a Dynamic User Equilibrium Assignment with Dynameq [19]. We would be interested in identifying which is the configuration of nodes that maximizes the total flow intercepted by the sensors located there; therefore, we formulate the objective function in terms of Rule 3 by representing the total intercepted flow. To enhance the OD covering, we impose the condition that at least one path of each OD pair has a detector located on it, and for practical reasons we also include a bounding constraint on the maximum number $\hat{l}$ of detectors that can be located in a network. The problem can then be formulated in terms of the following extended set covering model with side constraints:

$$\begin{aligned}
&\boldsymbol{MAX} \sum_{\boldsymbol{k \in K}} \boldsymbol{h_k \cdot y_k} \\
&s.t. \\
&\qquad \sum_{n \in N} x_n \leq \hat{l} \\
&\qquad \sum_{n \in N} \delta_{nk} \cdot x_n \geq y_k\,, \quad \forall k \in K_i, \forall i \in I \\
&\qquad \sum_{k \in K_i} y_k \geq 1, \quad \forall i \in I \\
&\qquad x_{n,} y_k \in \{0,1\}
\end{aligned} \tag{8}$$

In this formulation the problem can be infeasible if $\hat{l} < l_0$, where $l_0$ is the minimum number of detectors necessary to satisfy the OD covering constraints. Therefore we quantify the infeasibility when it appears. To this end, the proposed third intersection detection layout formulation is the previously presented formulation, adding the OD covering rule in a Lagrangian fashion. Let:

$$z_i = \begin{cases} 1 \; if \; i^{th} \; ODpair \; is \; covered \; at \; least \; by \; one \; detector \\ 0 \; otherwise \end{cases}$$

$$MAX\; F(y,z) = \alpha \cdot \frac{\sum_{k \in K} h_k \cdot y_k}{\sum_{k \in K} h_k} + \beta \cdot \frac{\sum_{i \in I} z_i}{|OD|}$$

$$\begin{aligned}
&s.t. \\
&\qquad \sum_{n \in N} x_n \leq \hat{l} \\
&\qquad \sum_{n \in N} \delta_{nk} \cdot x_n \geq y_k\,, \quad \forall k \in K_i, \forall i \in I \\
&\qquad \sum_{k \in K_i} y_k \geq z_i, \quad \forall i \in I \\
&\qquad x_{n,} y_k \in \{0,1\}
\end{aligned} \tag{9}$$

The sensitivity analysis of the Lagrangian multipliers $\alpha, \beta$ allows us to identify which is the total flow intercepted by the detectors that can be interpreted either in terms of an error bound or the quality of the traffic information that can be generated. The analysis of the infeasibility also reveals which are the uncovered OD pairs. A proper selection of $\hat{l}$ provides a solution that captures 100% of the traffic demand, ensuring in this way the complete observability of the system.

The location of Bluetooth sensors can raise an additional question in terms of the measurement of travel times between pairs of detectors along the likely used paths, as has been described in the introduction of this section. In order to achieve this objective, we propose a new formulation of the model that adds two sets of constraints:

- Ensuring a minimum number of detectors on each path, and
- Imposing a condition of minimum linear distance between two detectors. This constraint can also be justified from the technological point of view of minimizing the likelihood of improper detection due to signal overlapping.

The new formulation would then be:

$$
\begin{aligned}
& MAX \sum_{k \in K} h_k \cdot y_k \\
& s.t. \\
& \sum_{n \in N} x_n \leq \hat{l} \\
& \sum_{n \in N} \delta_{nk} \cdot x_n \geq p y_k\,, \quad \forall k \in K_i, \forall i \in I \;\; (*) \\
& \sum_{k \in K_i} y_k \geq 1, \qquad \forall i \in I \qquad\qquad (10) \\
& x_i + x_j \leq 1 \;\; \forall i, \forall j \in V(i) \\
& x_{n,} y_k \in \{0,1\}
\end{aligned}
$$

Where:

- $V(i)$ is the linear neighborhood of intersection i, defined as:
$$V(i) = \{j \in I | dist(i,j) \leq m\, meters\}$$
- $p$ is the parameter defining the minimum number of detectors per path, constraints (*) provide path information; $p > 1$ provides travel times between detectors along a path
- $m$ is the minimum linear distance between two detectors

## COMPUTATIONAL RESULTS FOR THE INTERSECTION LAYOUT MODEL

A set of computational experiments has been conducted in the same scenario proposed for the link covering problem: the Eixample district in Barcelona. An updated OD matrix corresponding to the traffic between 8 am and 9 am for a typical weekday of 2007 is utilized. The exact solutions for the different proposed experiments have been obtained with CPLEX. Due to space constraints we will report here only the computational results for models (9) and (10). Figure 4 depicts the sensitivity analysis of the Lagrangian multipliers in a restrictive case when the number of available detectors is fixed at 10 in model (9).

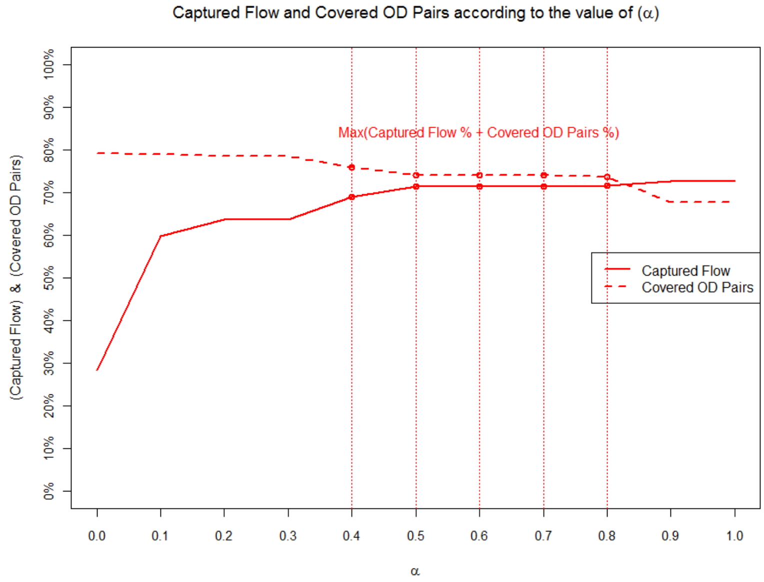


**Figure 4 - Lagrangian multipliers sensitivity analysis**

Table 3 summarizes the reverse experiment, fixing the Lagrangian multipliers $(\alpha, \beta)$ to 0.5, in order to simulate a situation where the relative importance of capturing flow and covering OD pairs is the same. Then, the model is executed with $\hat{l} = 1$ until $\hat{l} = 80$, increasing the value one by one.

| Used Detectors (D) | Captured Flow | Covered OD Pairs | O.F. |
|---|---|---|---|
| 1 | 12,65 | 17,66 | **15,16** |
| 2 | 21,97 | 31,58 | **26,77** |
| 3 | 30,53 | 43,74 | **37,14** |
| 4 | 40,29 | 48,89 | **44,59** |
| 5 | 45,38 | 56,02 | **50,70** |
| 6 | 50,13 | 61,29 | **55,71** |
| 7 | 56,23 | 65,15 | **60,69** |
| 8 | 62,04 | 69,12 | **65,58** |
| 9 | 66,87 | 71,58 | **69,22** |
| 10 | 71,50 | 74,15 | **72,83** |
| … | … | … | **…** |
| 15 | 84,52 | 85,26 | **84,89** |
| … | … | … | **…** |
| 20 | 91,90 | 92,28 | **92,09** |
| … | … | … | **…** |
| 25 | 95,59 | 96,49 | **96,04** |
| … | … | … | **…** |
| 30 | 98,26 | 98,83 | **98,55** |
| … | … | … | **…** |
| 40 | 99,83 | 100,00 | **99,91** |
| … | … | … | **…** |
| 80 | 99,98 | 100,00 | **99,99** |

**Table 3 – Summary of captured flow and covered OD pairs for $\alpha = \beta = 0.5$**

Figure 5 displays the graphics of the percentage of intercepted OD flows as a function of the number of detectors located by model (10).

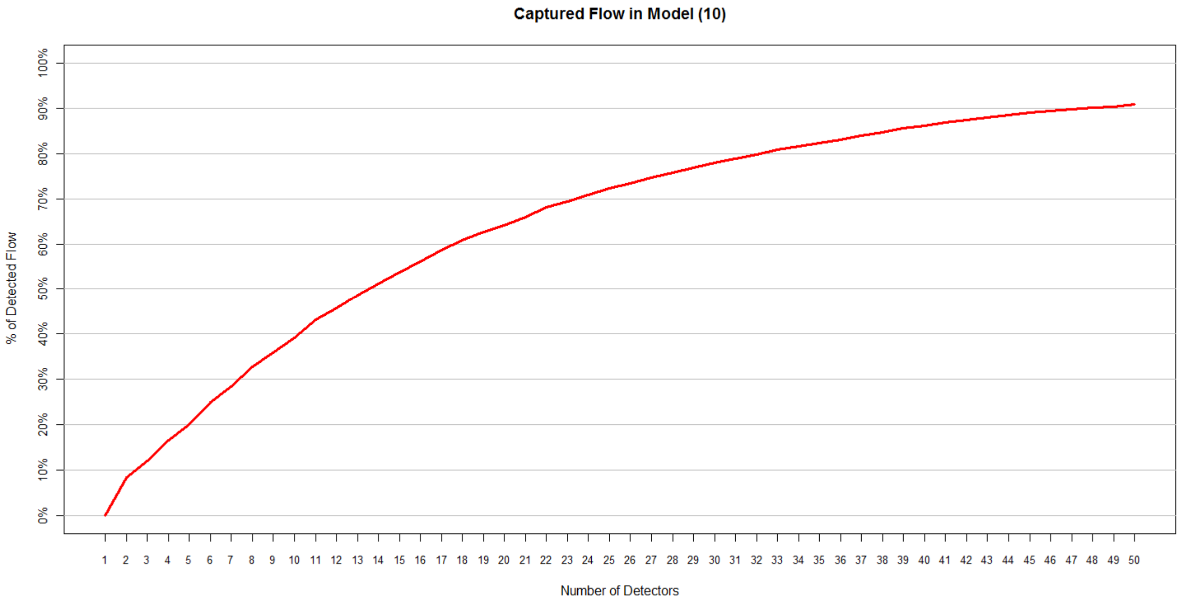
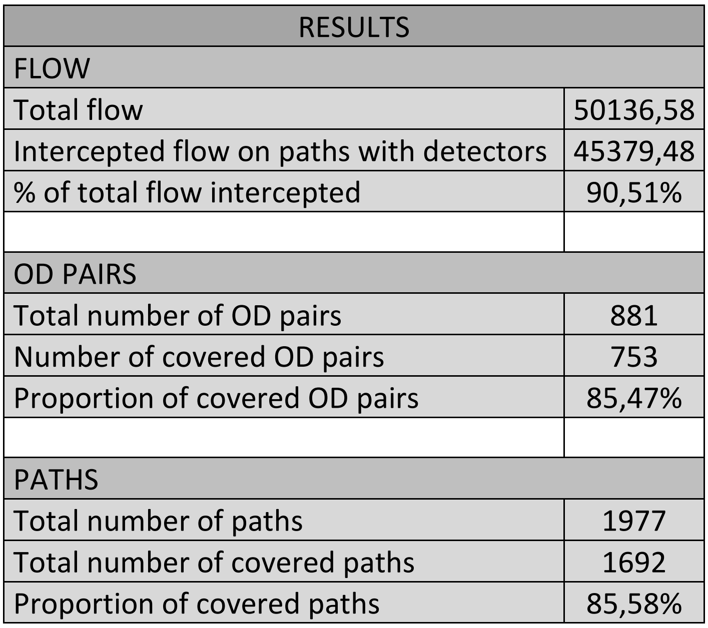


**Figure 5. Percentage of intercepted flow as a function of the number of detectors**



| RESULTS | |
|---|---|
| FLOW | |
| Total flow | 50136,58 |
| Intercepted flow on paths with detectors | 45379,48 |
| % of total flow intercepted | 90,51% |
| | |
| OD PAIRS | |
| Total number of OD pairs | 881 |
| Number of covered OD pairs | 753 |
| Proportion of covered OD pairs | 85,47% |
| | |
| PATHS | |
| Total number of paths | 1977 |
| Total number of covered paths | 1692 |
| Proportion of covered paths | 85,58% |

**Table 4 Summary of numerical results for model (10) for $\hat{l} = 50$ Bluetooth sensors**

Table 4 summarizes the analysis of the quality of the solution in terms of the total amount of intercepted OD flows and OD pairs for $\hat{l} = 50$, $p = 2$ and $m$=300 meters.

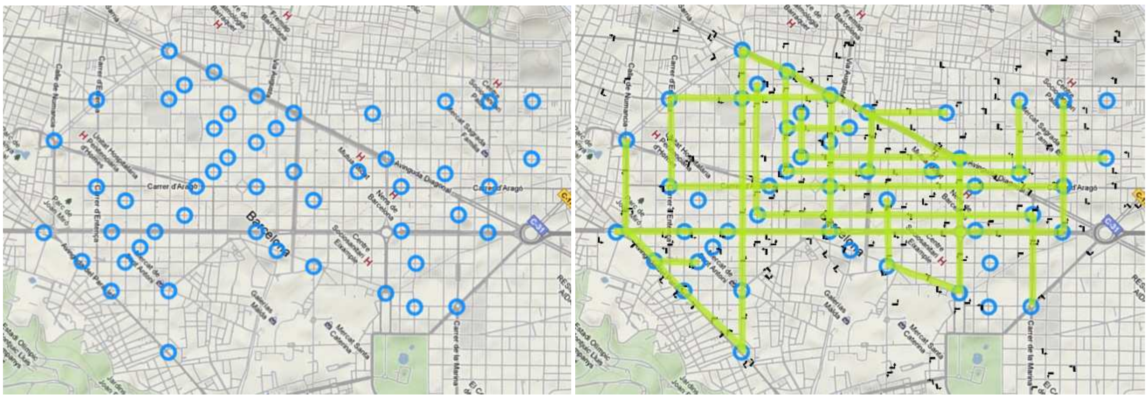


**Figure 6 Optimal location of 50 Bluetooth detectors by model (10) with *p*=2 and *m*=300**

Figure 6 displays the location of the 50 detectors and an example of some sections of the main paths whose travel times are estimated by Bluetooth detection.

## CONCLUSIONS

By considering only the most probable paths, the new approach proposed for the link detection layout problem reduces the size of the problem considerably when including time periods. Consequently, it ensures that a good estimation is possible. This formulation appeared to be very sensitive to pre-processing, which can further reduce the size of the problem. When solved with CPLEX, it appears to be well conditioned as almost no more pre-processing is found by the software, and the resolution time is very short. Even with time considerations, the number of detectors necessary to satisfy the OD covering rule stays reasonable, unlike other previous formulations with time periods. This shows that the new theoretical model is very good and realistic for practical use.

The reformulation of the problem in terms of intersection detection layout shows an even better performance in substantially reducing the number of detectors needed to maximize the total intercepted flow and/or the number of OD constraints.

## ACKNOWLEDGEMENTS

The research reported in this project has been funded by projects MITRA (TRA2009-14270 (subprogram MODAL, FEDER Co funded)) and In4Mo (TSI-020100-2010-690) of the Spanish R+D National Programs

## REFERENCES

[1] Florian M. and Hearn D. (1995), Network Equilibrium Models and Algorithms, Chapter 6 in: M.O. Ball et al., Eds., Handbooks in OR and MS, Vol.8, Elsevier Science B.V.

[2] Ben-Akiva, M., Bierlaire, M., Burton, D., Koutsopoulos, H.N. and Mishalani, R., (2001), Network State Estimation and Prediction for Real-Time Traffic Management, Networks and Spatial Economics, 1, pp. 293-318.

[3] Cascetta, E. (2001) Transportation Systems Engineering: Theory and Methods, Chapter 8, Kluwer Academic Publishers.

[4] Van Zuylen, H.J. and Willumsen, L.G. (1980) The most likely trip matrix estimated from traffic counts, Transportation research Part B, 14, pp. 281-293.

[5] Codina E., Barceló,J. (2004) Adjustment of O-D matrices from observed volumes: an algorithmic approach based on conjugate gradients, European Journal of Operations Research, Vol. 155, pp. 535-557.

[6] Lundgren J. T. and Peterson A. (2008), A heuristic for the bilevel origin–destination-matrix estimation problem, Transportation Research Part B 42 (2008) 339–354

[7] Ashok, K. and Ben-Akiva, M., (2000) Alternative Approaches for Real-Time Estimation and Prediction of Time-Dependent Origin-Destination Flows, Transportation Science 34(1), 21-36.

[8] Antoniou, C., Ben-Akiva, M. and Koutsopoulos H.N., (2007), Nonlinear Kalman Filtering Algorithms for On-Line Calibration of Dynamic Traffic Assignment Models, IEEE Transactions on Intelligent Transportation Systems Vol. 8 No.4.

[9] Barceló J., Montero L., Marqués L. and Carmona C., (2010), Travel time forecasting and dynamic of estimation in freeways based on Bluetooth traffic monitoring, Transportation Research Records: Journal of the Transportation Research Board, Vol. 2175, pp. 19-27.

[10] Bierlaire, M., (2002), The Total Demand Scale: a New Measure of Quality for Static andDynamic Origin-Destination Trip Tables, Transpn. Res. B 36, 837-850.

[11] Castillo, E. Conejo, A. J. Menéndez, J. M. Jiménez, P., (2008, The Observability Problem in Traffic Network Models, *Computer-Aided Civil and Infrastructure Engineering* 23, 208–222

[12] Yang, H., and Zhou, J. (1998), “Optimal Traffic Counting Locations for Origin-Destination Matrix Estimation”, Transp. Res. B, Vol. 32B, No.2, pp.109-126

[13] Larsson, T. Lundgren, J.T. and Peterson, A. (2010), Allocation of Link Flow Detectors for Origin-Destination Matrix estimation a Comparative Study, Computer-Aided Civil and Infrastructure Engineering, 25, pp. 116-131.

[14] Ehlert, A.,. Bell, M. G.H and Grosso, S. (2006), The Optimisation of Traffic Count Locations in Road Networks, Transportation research Part B 40, pp. 460-479.

[15] Fei, X., Eisenman, S. M. and Mahmassani, H.,(2007) Sensor Coverage and Location for Real-Time Traffic Prediction in Large-Scale Networks, Paper presentes at the 86th Annual Meeting of the Transportation Research Board, January 2007.

[16] Hoogland, J.K. (2000), The traffic sensor location problem, Master Thesis, Delft University

[17] 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 35, pp. 50-60.

[18] Larsson, T., Lundgren, J.T., Patriksson, M. and Rydergreen, C. (1998), Most likely equilibrium route flows: analysis and computation, Proceedings of TRISTAN III.

[19] Dynameq 1.4, (2010), INRO Consultants, Montreal

[20] Savelsbergh, M.W.P. (1994), Preprocessing and Probing Techniques for Mixed Integer Programming Problems. ORSA J. Comput. 6, 445-455.

[21] Hertz, A., Taillard, E.D., de Werra, D., (1997) A Tutorial on Tabu Search. Local Search in Combinatorial Optimization, E. Aarts and J.K. Lenstra (Eds), 121-136.

[22] Glover, F. (1986) Future Paths for Integer Programming and Links to Artificial Intelligence. Computers and Operations Research 13, 533-549.