# Cellular Flow in Mobility Networks

Alfredo Milani, Eleonora Gentili and Valentina Poggioni

***Abstract*—Nearly all the members of adult population in major developed countries transport a GSM/UMTS mobile terminal which, besides its communication purpose, can be seen as a mobility sensor, i.e. an electronic individual tag. The temporal and spatial movements of these mobile tags being recorded allows their flows to be analyzed without placing costly ad hoc sensors and represents a great potential for road traffic analysis, forecasting, real time monitoring and, ultimately, for the analysis and the detection of events and processes besides the traffic domain as well. In this paper a model which integrates mobility constraints with cellular networks data flow is proposed in order to infer the flow of users in the underlying mobility infrastructure. An adaptive flow estimation technique is used to refine the flow analysis when the complexity of the mobility network increases. The inference process uses anonymized temporal series of cell handovers which meet privacy and scalability requirements. The integrated model has been successfully experimented in the domain of car accident detection.**

***Index Terms*—Mobile networks, spatial data mining, traffic flow analysis.**

## Introduction and Related Work

THE basic laws governing human mobility are becoming an essential part in scientific works ranging from urban planning, road traffic forecasting to spread of biological viruses [1], contextual marketing and advertising. New opportunities arise for the study of human mobility with the advent of the massive diffusion of mobile networks for personal devices such as GSM, UMTS, IEEE 802.16 WiMAX, IEEE 802.11 WLAN. Nearly all the members of adult population in major developed countries transport a GSM/UMTS mobile terminal, i.e. an electronic individual tag, with themselves. Moreover, in order to provide the service, the data of GSM/UMTS networks are already logged by mobile phone companies. The analysis of temporal and spatial movements of these mobile tags allows accurate estimation of urban/extraurban traffic flow without placing costly ad hoc sensors. Mobile network data represent a powerful mean for road traffic analysis, forecasting and real time monitoring and, ultimately, for the analysis and the detection of events and processes besides the traffic domain(e.g. traffic jam, velocity, congestions, road work, accidents etc.), which can affect the motion behavior of the masses (e.g. sport and leisure events, concerts, attractive shopping areas, working/living area cyclic processes etc.).

Techniques and models for mobile device flow analysis [2] have mostly focused on predictive models aiming at optimizing some mobile network system parameters such as cell dimensioning, antenna distribution, and load balancing [3], [4].

Alfredo Milani, Eleonora Gentili, Valentina Poggioni are with the Department of Mathematics and Computer Science, University of Perugia, Italy.

On the other hand a number of projects [3], [5] try to use the cellular network traffic to estimate different road traffic and transportation related quantities [6], [7], [8], such as speed and travel times between destinations [9], [10], [11], origin/destination (O/D) matrices [12], [2], road traffic congestions [11], road traffic volume or density [13], [14], etc.

Many projects are also active in the relatively recent area of mobile device localization which focuses on the position of the single mobile terminal for the purpose of providing spatial contextual services.

The main limitation of the existing approaches to traffic estimation is the lack of a model taking explicitly into account of the mobility and transportation infrastructures. The estimates are often based on purely statistical correlation approaches which usually assume users movement directions following a uniform probability distribution. On the other hand, physical and normative constraints to user mobility inside a cell (e.g. as roads topology, mandatory directions etc.) are usually not taken into account in those models, with few exceptions [15], [4], while relationships with traffic domain external events, such as social events and social processes (e.g. work/home commuting, shopping periods etc.) are completely ignored.

Moreover some issues such as *privacy* and *scalability* are also problematic. For instance, techniques for inferring $O/D$ matrices [2] uses information about the *Location Areas (LA)* over the time, where a $LA$ is a set of cells where the mobile terminal is assumed to be located. In other words the algorithm needs to identify time, origin and destination $LA$s of the whole trip made by each single telephone, thus representing a remarkable privacy infringement. Mobile device localization detect the spatial position of the single user, by using techniques based on distance from the cell antenna (for example in [2]), or assuming the placement of special detector antennas for enhancing the accuracy of the localization. Although the remarkable precision is obtained, in both cases there are relevant problems of privacy and scalability. In fact, due to the huge amount of data generated by monitoring, each single terminal position in a cell would requires an enormous bandwidth, storage and computational cost.

In this work we propose a model which integrates spatial networks with mobile phone networks, in order to monitor, analyze and predict the user traffic on the mobility infrastructure and to make detection and inference about social events and processes in place, on the basis of anonymous aggregated data. The aim is that by integrating mobility constraints (e.g. available roads), it is possible to improve the accuracy of predictions the cellular network based on the mobility/transportation network and vice versa. Moreover social event/processes which take place can also be detected, and conversely the knowledge of those events/processes can improve the predictive model in the mobility domain.

In particular consider temporal data series describing the ”handover” of anonymous users, i.e. the number of users which traverses any of the six boundaries of an hexagonal cell in a mobile phone network. The choice of handover data is due to different reasons: (1) *Privacy issues*. Anonymized handovers can easily be made available and can be securely and effectively transmitted while tracking the positions of a single terminal would represent sensitive data about the individual user behavior. (2) *Performance* and *scalability issues*. The size of the information to process remains constant as the number of users increase.

The rest of the paper is organized as follows: Definitions and relationships between spatial networks and mobility networks are introduced in Section I, while a model for inferring spatial mobility flow from one word data is presented in Section II. An adaptive estimate model, used as a basis for event detection, is presented in Section III. Experiments for car accident detection are presented in Section IV, and discussion on possible directions for future works in Section V concludes the article.

## I. Spatial network and Cellular Network

In this section we introduce a model for integrating the knowledge of a spatial network which constraints users movement and the knowledge of the cellular network covering the same physical area.

A *spatial network*, or *mobility network*, is a set of physical means and normative constraints, such as roads, railways, underground transportation, pedestrian area, one-way lanes and highways, which narrow the mobile user mobility. In general more than one cellular network with different sizes and topologies can insist on the same area. Here we assume that a single *cellular network* is operating in the given area and it is organized in the usual hexagonal grid of cells with each antenna centered in a cell. According to the usual notation, given a reference cell, say cell 0, we refer to its neighbour cells, by numbers from 1 to 6 clockwise as shown in Fig.1.

A *spatial network* $S$ can be described as $S = (N, A, D, loc)$ where $N$ are nodes, $A \subseteq N \times N$ are directed arcs, and $loc$ is a *location* function $loc : N \to D$ mapping nodes onto positions in the bi-dimensional area of interest $D \subseteq \mathbb{R}^2$.

A *cellular network* $M = (C, D, g, m)$, organized in an hexagonal grid, is defined by a set of cells $C$, a function $g : C \times \{1 \ldots 6\} \to C$, which describes the *grid topology*, ($g$ returns the i-th neighbour of a given cell or returns the same cell if no i-th neighbour exists,e.g.it is on the border) and a boolean function $m : C \times D \to \{T, F\}$ checking whether a given position of $D \subseteq \mathbb{R}^2$ belongs to a cell.
Note that $g$ and $m$ should verify the hexagonal grid topology.

When a cellular network $M$ shares the same domain area $D$ with a spatial network $S_0$, we can consider the spatial network ”projection” over the cells, or equivalently we can see $C$ as ”cutting” $S_0$ into a family of disjunct spatial subnetworks $\{S_i\}$. In order to identify the spatial subnetwork corresponding to each cell, it is useful to introduce additional nodes in correspondence of the cutting edges whenever an arc of the spatial network crosses the boundary between two cells.

*Projected spatial network*. The projection $S = \pi(M, S_0)$ of $S_0 = (N_0, A_0, D, loc_0)$ according to a cellular network $M = (C, D, g, m)$, is the spatial network $S = (N, A, D, loc)$ obtained by $S_0$, such that
1) $\forall n \in N_0 \Rightarrow n \in N$ and $loc_0(n) = loc(n)$,
2) $\forall (n', n'') \in A_0$ s. t. $\exists c \in C$ with $m(loc(n'), c) = m(loc(n''), c) = T$ then $n', n'' \in A$, i.e. all the arcs in $S_0$ which originates and ends in the same cell, also belong to $S$,
3) for each arc $(n', n'')$ whose ends do not lie in the same cell, let $m(loc(n'), c') = m(loc(n''), c'') = T$ such that $c' \neq c''$ are two neighbor cells, then a new node $n'''$ and two new arcs, respectively $(n', n''')$ and $(n''', n'')$ will be added to the set of network nodes $N$ and arcs $A$; the position of the new node $loc(n''')$, will be assigned such that the node lies on the border between the two neighbors cells (note that $n'''$ belongs to both cells, i.e. $m(loc(n'''), c')$ and $m(loc(n'''), c'')$ are both true ),
4) finally, if an arc of $S_0$ traverses more than two cells, then the arc is cut in a series of subarcs according to the previous procedure.

An example of a spatial network and its projection on a cellular network is shown in Fig.1.

*Cell spatial network*. The projection operation $\pi(M, S_0)$ partitions $S$ into subnetworks. In particular for each cell $c \in C$ there is an associated *cell spatial subnetwor* $S \mid_c$ defined by the restriction of $S$ to all nodes and arcs lying inside $c$, i.e. in the domain area $D \mid_c = \{d \in D \mid m(c) \text{ is true}\}$. It is possible to identify in $S \mid_c$ two family of sets of nodes $I_{c,c_i} \subseteq N$ ( respectively $O_{c,c_i} \subseteq$ N) for $i = 1 \ldots 6$, which represent the set of nodes on the edge between the neighbors $c_i$ of the cell $c$ and connect inbound (outbound) arcs of $c$ with outbound (inbound) arcs of $c_i$. The set of nodes $I_c = \bigcup_{i=1}^{6} I_{c,c_i}$ and $O_c = \bigcup_{i=1}^{6} O_{c,c_i}$ represent respectively the *source* and *sink* nodes for the spatial subnetwork limited by cell $c$. Since after the projection, by construction, it does not exists any arc of $S$ crossing cell boundaries, $I_c$ and $O_c$ are the only sources and sinks for the flow in cell $c$.

The projection operation $\pi$ is defined by successive incremental splits upon properties of connectivity of the spatial graph and the cell area domains. It is easy to see that projection process can be extended for more complex characterizations of the spatial network which consider features on arcs or nodes, such as costs, distances, speed and time between nodes, capacities and probabilities.

## II. An Integrated Model for Spatial and Communication Network

Given a spatial network $S \mid_c$ delimited by a given cell $c$ (cell 0 or $c_0$ in the following) the amount of user flow inside/outside the cell is completely described by the data available from the cell control unit. Assume that $U_0^t$ denotes the amount of users in the current cell $c$ at the time slot $t$(stationary users); $HO^t(i, j)$ represents the handovers, i.e. the amount of mobile terminals moving from the cell $c_i$ towards the cell $c_j$ at the time $t$, then $HO_{in}^t = \sum_{i=1}^{6} HO^t(i, 0)$ and $HO_{out}^t = \sum_{i=1}^{6} HO^t(0, i)$ represent respectively all the users coming in and going out the reference cell at time $t$. In order to relate these data to the traffic flow in the different parts the mobility network we need to introduce some definitions.

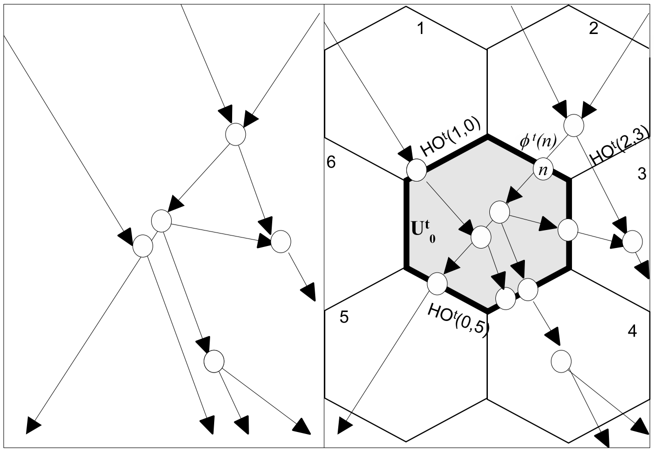


Fig. 1. Spatial mobility network and cellular network projection

Let $P$ the set of connected components of $S\mid_c$, for each $p \in P$:

- $IN(p)$ is the set of the *source nodes* of the component $p$, i.e. all the nodes in $p \cap I_{c,c_j}$, $\forall j = 1..6$,
- $OUT(p)$ is the set of the *sink nodes* of the component $p$, i.e. all the nodes in $p \cap O_{c,c_j}$, $\forall j = 1..6$.

Moreover $ON(k)$ is the set of all the nodes in $IN(p)$ and $OUT(p)$ lying on the edge between cell $c$ and its neighbor $c_k$.

### A. Mobility Network Flow Equations

Given the $HO$ series between the current cell $c$ and all its neighbors $c_1 \dots c_6$, and given the network topology $S \mid c$ projected on the cell 0, it is possible to define an inference model for deriving the flow of mobile users on the mobility network.

The model is based on the flow equations which relate the user flow in the cellular network with the flow in the spatial network that restricts user mobility. Let $U_0^t$ be the amount of users in the current cell $c$ and $HO_{in}^t$, $HO_{out}^t$ the handover data at the time-slot $t$. The flow on the spatial network delimited by the cell $c$ is *admissible* if

$$HO_{in}^t + U_0^t = HO_{out}^{t+1} + U_0^{t+1}. \tag{1}$$

Considering the set $P$ of all connected components, we can assume that for any $p \in P$ there exists an admissible flow, and let $\phi^t : N \rightarrow \mathbb{N}$ be the function assigning to each node $n \in N$ the number of users $\phi^t(n)$ in the node $n$ at the time $t$ and $U_{0,p}^t$ the users stationary at the time $t$ in the nodes of $p$ inside the cell 0, then

$$\forall p \in P, \sum_{n \in IN(p)} \phi^t(n) + U_{0,p}^t = \sum_{n \in OUT(p)} \phi^{t+1}(n) + U_{0,p}^{t+1}. \tag{2}$$

Assuming that only the handover and stationary users data are available from the cellular network, it is not possible to know how the users are distributed over the paths in the cell. So for each connected component $p \in P$ we know the exact values of $\phi^t(n)$ and $U_{0,p}^t$ only when, for each edge $k$ of cell 0, $IN(p) \cap ON(k) = \{n_1\}$ and $OUT(p) \cap ON(k) = \{n_2\}$. In this case we have $\phi^t(n_1) = HO^t(k,0)$ and $\phi^t(n_2) = HO^t(0,k)$.

Let consider the following equivalence relation $\sim$ between the elements of $P$: $\forall p_1, p_2 \in SP$ then $p_1 \sim p_2 \Leftrightarrow \exists o_1 \in OUT(p_1)$, $\exists o_2 \in OUT(p_2)$, $\exists k_1 \in \{1,\dots,6\}$ : $o_1, o_2 \in ON(k_1)$, or $\exists i_1 \in IN(p_1)$, $\exists i_2 \in IN(p_2)$, $\exists k_2 \in \{1,\dots,6\} : i_1, i_2 \in ON(k_2)$.

The relation $\sim$ partitions $P$ into equivalence classes having either sources or sinks on the same side of the cell. Therefore it can be more useful to provide (2) with respect to the $\sim$ equivalent classes. Since the connected components having paths on the same edge of the cell belong to the same equivalent class, and since

$$\sum_{n \in IN(p) \cap ON(k)} \phi^t(n) = HO^t(k,0),$$

$$\sum_{n \in OUT(p) \cap ON(k)} \phi^t(n) = HO^t(0,k),$$

both equations (1) and (2) can be rewritten for each equivalence class induced by the relation $\sim$.

In practice, the equivalent classes can be thought as *clusters of paths* originating from or sinking to the same set of cells.

Fig.2 represents some possible spatial networks related to the reference cell. In Fig.2.a only one connected component exists. Then, the general flow equation (1) coincides with the one of the connected component. In this case our model is exact to estimate the number of users in the paths and we say that we reach *component level* accuracy. In Fig.2.b we can see two connected components belonging to two different clusters. In this case we reach *component level* accuracy. The cases represented in Figs.2.c and 2.d are equivalent in terms of handover data, but they are different from the topological point of view. While Fig.2.d has a unique connected component, we have two connected components in Fig.2.c which belong to the same equivalent class. Even if an equation for each connected component can be written, the handover data are provided for each edge (and not for single path). So the accuracy level decreases to *cluster level*.

### B. Inferring user flow

Assuming that the initial number of mobile user in component cluster $c$ at the initial time slot 0, is known, it is possible to infer the number of stationary users in a given time slot in the cluster by iteratively applying the flow equations generated by the spatial network on a cell $C$.

In fact, from the general equation of admissible flow (1), for each cluster of components (i.e. for each equivalent class) we have:

$$U_0^{t+1} = HO_{in}^t + U_0^t - HO_{out}^{t+1}.$$

With consecutive substitutions, we obtain

$$U_0^{t+1} = HO_{in}^t + HO_{in}^{t-1} + U_0^{t-1} - HO_{out}^t - HO_{out}^{t+1};$$

By regroupings terms, we obtain

$$U_0^{t+1} = \sum_{j=0}^{t} HO_{in}^j + U_0^0 - \sum_{j=0}^{t} HO_{out}^{j+1}, \tag{3}$$

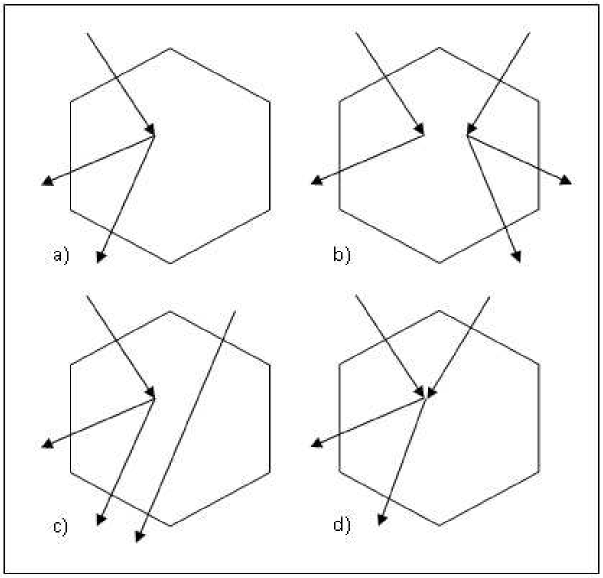


Fig. 2. Connected components and clusters in cell spatial network

.

The inference procedure assumes that the amount of users in stationary state inside every cluster at time 0 is known.

It is easy to note that the ability of distinguish the flow within a connected component of $S\ |_c$ is limited by the number of classes induced by $\sim$. In other words if two connected components are in the same class the amount of their individual inbound/outbound flow cannot be precisely determined considering only the cell handovers. In the ideal case if each connected component belongs to a distinct class, its flow is fully described by the handovers data.

The effectiveness and accuracy of the inference technique, based solely on handover data, greatly depends on the cell resolution/granularity, i.e. the relative size of the cell with respect to the spatial network, and on the spatial network connectivity. For instance the presence of high connectivity subnetworks or hubs, such a square or a park, where the mobile phone holders can move "freely" in any direction, can narrow down the accuracy. On the other hand, a cell covering an highway section in an area with no other road can provide high accuracy.

## III. Spatial Networks Prediction and Estimation

In this section we present a prediction model based on Markov chain and an adaptive flow estimation model which exploits the underlying spatial network in order. These models can improve the performance of predictions and give a better estimate the flow within the single connected component when inference based on flow equations cannot determine a unique answer, i.e. the equations have not a unique solution due to large clusters of components. The two models represent the core modules of the event detection system presented in the Experiments Section.

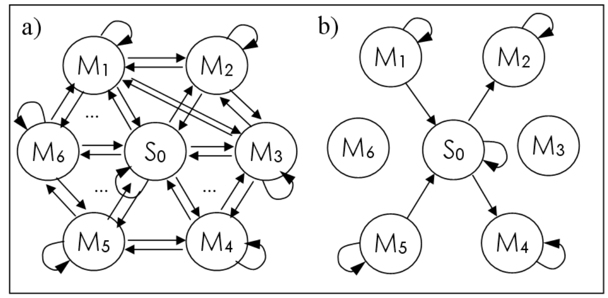


Fig. 3. State diagram: a)complete b)reduced by mobility constraints

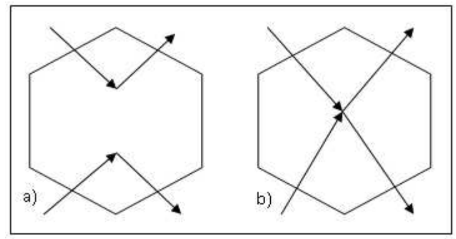


Fig. 4. An example of different spatial networks that are unrecognizable from their handover

### A. Prediction Markov Model

The prediction model is based on the Markov model proposed in [4] for mobile network management. Mobile user movements towards/from the cell are represented by a state diagram associated to a transition matrix assigning probabilities assigned to each movement in the given time slot. A complete state diagram for 7 direction levels is represented in Fig.3. The parameters of the Markov model, i.e. the specific transition probabilities, can be effectively determined by a statistical analysis of handover series at the given time slot granularity.

It is worth noticing that spatial network constraints can reduce the number of states, the entries of the incidence matrix and thus the complexity of the Markov model. For instance the projected spatial network of Fig.4.a can reduce the predictive model to 4 states as shown in Fig.3.b. Nevertheless the Markov model is not adequate by itself for flow analysis since qualitatively different mobility networks, as the ones shown in Fig.4 can lead to the same Markov model structure. The technique shown in Section II-B can be applied in order to calculate the flow in clusters of connected components.

### B. Flow estimation

In order to improve the accuracy of flow inference within a class of connected component, it is possible to use an estimate of flow distributions on sources/sinks, when deterministic inference is not possible.

Assume that for each set of source nodes $I(c, c_i)$ (sink nodes $I(c, c_i)$) lying on the same border $i$ of cell $c$, the distribution $\rho^t(n)\ \forall n \in I(c, c_i)$ i.e. the expected percentage of handovers $H(i, 0)$ ($HO(0, i)$ )which take place at time $t$ because of users entering (leaving) cell $c$ from node $n$ is known. It is apparent that the flow equations can be restated in the form of estimate for each single connected component $\forall p \in P$,

$$\sum_{\substack{n \in IN(p) \\ k \in 1 \ldots 6}} \sigma^t(k,0) + V^t_{0,p} = \sum_{\substack{n \in OUT(p) \\ k \in 1 \ldots 6}} \sigma^{t+1}(k,0) + V^{t+1}_{0,p} \quad (4)$$

where the term $\sigma^t(k,0) = HO^t(k,0) * \rho^t(n)$ represents the estimated flow through each source/sink node and $V$ the current estimated stationary users calculated iteratively. The estimate can also be propagated along the mobility network and between cells by simple boundary equations, since incoming flow for a cells is the outgoing flow for its neighbor and vice versa.

It is worth noticing that the flow estimates are also required to be *admissible*, i.e. they should not contradict the general flow equations. On the other hand, contradictions can emerge over the time by iterating wrong estimates. For example a low estimate of flow distribution along a path can lead to observing many more users than expected exiting from that path in a neighbor cell. In this case, for example, the distribution parameters can be increased along the contradictory path to re-establish the consistency.

On this basis it is possible to design a scheme for adaptive flow estimation, where the estimation parameters are dynamically changed in order to maintain consistency between the estimate and the observed data (i.e. handover and total stationary users):

*Adaptive Flow Estimation Scheme*:

1) Estimate current flow along sources and network paths of cell $c$ using the real data and current distribution parameters
2) Calculate flow constraints in neighbor cells of $c$ using current estimate data and parameters
3) If current estimate conflicts with the previous constraints for cell $c$ then (3.1) revise distribution parameters, $\sigma^{t+1} = r(\sigma^t)$ to establish consistency and (3.2) back-propagate revision to $c$ neighbors.

A key point of the adaptive algorithm is the update function $r$ which revises the estimate distribution parameters $(\sigma^t(k,0))$. The current implementation uses an iterative algorithm based on PSO [16]–[18] to find the increment/decrement size distribution for re-establishing the consistency.

## IV. Experiments: Car accident Detection in Highways

The proposed model for the analysis and the estimation of traffic flow in mobility network has been experimented in the domain of car accident detection in the Great Ring Highway (GRH) A90 surrounding the city of Rome in Italy. Timed data series of handover logs from a major national GSM mobile phone network has been used. The provided data regard 32 months for a cluster of 24 GSM cells covering a section of the GRH with different cells dimensions and road density in the domain area (see Fig.5). In addition, reports from the national highway traffic control system have been used as a source of car accidents events in the GRH; the salient types of information include: *start/end time of event (i.e. return to normal traffic condition), place and direction of the event, class of traffic impact* (from 0=null to 6=complete block). The event features which have been considered are: *start/ending time*, *place* (i.e. mobility network connected component), *direction* (which GRH lane is concerned for the event), and *type of event* (car accident or generic anomalous events).

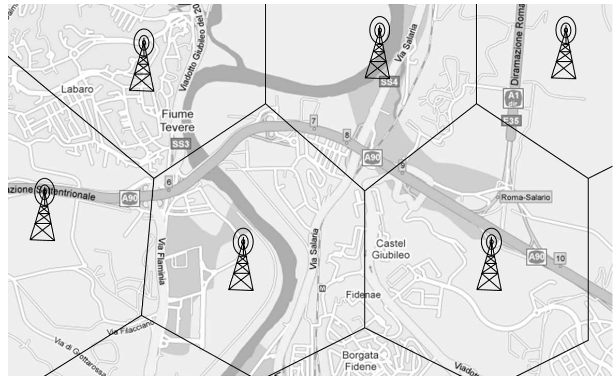


Fig. 5.

The data of the first 24 months have been used to determine the initial values for the adaptive estimation model and the weights of the Markov predictive model and alert thresholds, while data of the last six months have been used from the actual experiment of detection. The parameters have been computed for each 15 minutes time slot on a week day base, Monday to Saturday, while Sundays and public holidays have been included in a different class, since their traffic behaviors exhibit common similar patterns.

The general architecture of the detection systems is based on different classes of indicators and thresholds which trigger alerts in the algorithm. Indicators based on global handover traffic in the cell are compared with the predictive model in order to detect start/end and type of events, while indicators of deviation from the adaptive estimation model are used to detect the place of the event the direction of accident. The scheme for the event detection loop is depicted below:

**if** event($HO_{in}$,$HO_{out}$) **then**
  eventStarted $\leftarrow$ true
  **if** carAcc($HO_{in}$,$HO_{out}$) **then**
    **if** carConn() **then**
      output estimatePlace()
      output estimateDirection()
    **end if**
  **end if**
**else**
  eventStarted $\leftarrow$ false
**end if**

Any start/end event is firstly detected by a relevant change in global handover volume $HO_{in} + HO_{out}$ with respect to the expectation according to the Markov based model. The value of the corresponding threshold $\vartheta^t_g$ is based on the variance of handovers volume ($g$ represents the event type). The beginning of an event of type car accident is related with a sudden increase of the number of $HO_{in}$ with respect to $HO_{out}$, see Fig.6). A threshold $\vartheta^t_{car}$ is compared against the averaged difference $HO_{in} - HO_{out}$ over consecutive time

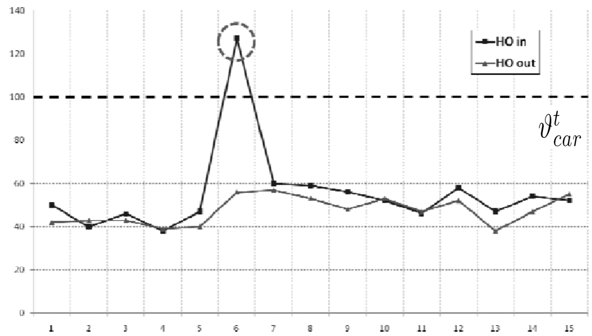


Fig. 6. A peak in incoming handovers over threshold $\vartheta^t_{car}$.

TABLE I
DETECTION RESULTS

| | Events | Precision | Start $\leq$ | Direction |
|---|---|---|---|---|
| $Alg_{est}$ | 382 | 97 % | 85 % | 98 % |
| $Alg_{det}$ | 495 | 75 % | 83 % | 65 % |

intervals to distinguish car accidents from other events such as anomalous increment of traffic. On the other hand the "end" of the event is recognized by a return to a normal traffic condition. When a potential car accident is detected, the estimated flow allows ones to determine the location of the event, which is decided to be the connected component/s with the greater estimated flow variation. The direction of the event is calculated by comparing the inbound/outbound estimated flow in the connected component corresponding to the highway lanes. The control *carConn()* filters out those flow variations which are due to car accidents already detected in the nearby cells and could be erroneously recognized as new events. The experimental results of the original algorithm $Alg_{est}$ have been compared with a version, $Alg_{det}$, which does not take into account of mobility network estimates, but only uses the deterministic inference rules. Car accidents with null effect on the traffic have been excluded from the statistics.

As shown in Table I, the results are quite encouraging: Both $Alg_{est}$ and $Alg_{det}$ algorithms detected all the 371 accidents events in the traffic control report, while $Alg_{det}$ has a considerable lower precision with a remarkable number of false positives (*Events* and *Precision*). It is interesting to note that the *start of event time* returned by both algorithms (*Start*$\leq$) is better, i.e. anticipated, with respect to the starting time given by the national traffic control system. This is because the mobile users data are acquired in real time, while the accident alert reach the national traffic system by different channels, e.g. drivers, police patrols etc., which are not always promptly activated.

The number of false positives (*Precision*) of $Alg_{det}$ is mostly due to the inability of distinguishing the "noise" of events taking place in the urban area nearby the highway, while $Alg_{est}$ uses the analysis of the traffic on the urban connect component to filter out events not taking place in GRH. The accuracy of direction detection (*Direction*) is found to be high. Failure of detection are sometimes inevitable due to a number of reasons. For example, car accidents in a lane sometimes can slow down the traffic in the other one for different reasons: rescuing cars blocking it, traffic police deviating the traffic on the other lane or the phenomenon of "accident curiosity" which draws the attention of drivers on the event slowing down the opposite lane traffic. If this happens within the first 15 minutes time slot, the algorithm is not able to detect a suitable direction since the two cannot be distinguished, while a finer time granularity in the data is expected to improve the direction detection ability. A further analysis has shown that most of the false positives detected by $Alg_{est}$ are due to traffic variation induced by car accidents in nearby cells. This suggests that the management of connected events should be further refined.

## V. FUTURE DEVELOPMENTS AND CONCLUSIONS

Cellular networks, besides their communication purpose, can be seen as mobility sensor networks already in place which offer a great potential for the analysis of users flow in an area. A model which integrates mobility constraints and cellular networks has been proposed in order to analyze, monitor, forecast and detect events and processes in the mobility infrastructure. The use of cell level handover meets data privacy and scalability requirements, while the knowledge of the mobility infrastructure allows ones to obtain reasonable estimates of the flow at the connected component level. The integrated model and the proposed technique of adaptive flow estimation have been successfully experimented in the domain of car accident detection.

Future works includes the investigation of techniques for the application of the model to high density urban area, where the high road density does not allow a fine grain analysis of the flows, although the increasing diffusion of the so called microcells and nanocells is soon expected to provide a suitable granularity.

More generally suitable models, which integrate "sensors already in place" (e.g. cellular networks, payment systems, bus/train ticketing systems, video surveillance etc.) and mobility infrastructures constraints, are of great interest for the analysis of social events (e.g. entertainment, sport events, festival, commercial/leisure area attractors etc.) and social processes (e.g. working day/vacation days cycle, work/school/home cycle etc.) which involve movement of people in the physical space and conversely, for analyzing the impact of events on the mobility infrastructures and their planning and management.

## REFERENCES

[1] M. C. Gonzales, C. A. Hildago, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453 (5), pp. 779–782, 2008.

[2] N. Caceres, J. Wideberg, and F. Benitez, "Deriving traffic data from a cellular network," in *13th World Congress for ITS system and services*, vol. 3, 2006, pp. 1–10.

[3] ——, "Review of traffic data estimations extracted from cellular networks," *Intelligent Transport Systems, IET*, vol. 2, no. 3, pp. 179–192, Sept. 2008.

[4] P. Fülöp, K. Lendvai, T. Szálka, S. Imre, and S. Szabó, "Accurate mobility modeling and location prediction based on pattern analysis of handover series in mobile networks," in *MoMM '08: Proceedings of the 6th International Conference on Advances in Mobile Computing and Multimedia*. New York, NY, USA: ACM, 2008, pp. 219–226.

[5] D. Valerio, A. D'Alconzo, F. Ricciato, and W. Wiedermann, "Exploiting cellular networks for a road traffic estimation: a survey ad a research roadmap," in *IEEE 69th Vehicular Technology Conference (IEEE VTC '09-Spring)*, Barcelona, Spain, April 26-29 2009.

[6] N. P. 70-01, "Probe-based traffic monitoring: State of the practice report," University of Virgina - Center for Transportation Studies, Tech. Rep., November 2005.

[7] J. Virtanen, "Mobile phones as probes in travel time monitoring," Finnish Road Administration, Helsinki, Finland, Tech. Rep., 2002.

[8] M. Höpfner, K. Lemmer, and I. Ehrenpfordt, "Cellular data for traffic management," in *6th European Congress and Exibition on Intelligent Transport System and Services*, 2007.

[9] Y. Youngbin, "The state of cellular probes," California PATH Research Project, University of California, CA, USA, Tech. Rep., July 2003.

[10] J. Remy, "Computing travel time-estimates from gsm signalling messages: the strips project," *Intelligent Transport Systems, IET*, 2001.

[11] J. Ygnace, "Travel timespeed estimates on the french rhone comdor network using cellular phones as probe," INRETS, Lyon, France, Lyon, France, Tech. Rep., December 2001.

[12] J. White and I. Wells, "Extracting origin destination information from mobile phone data," in *11th Int. Conf. on Road Transport Information and Control, LONDON*, 2002, pp. 30–34.

[13] C. Ratti, R. Pulselli, S. Williams, and D. Frenchman, "Mobile landscapes: using location data from cell-phones for urban analysis," *Environ. Plann. B Plann. Des.*, vol. 33 (5), pp. 727–748, 2006.

[14] K. THIESSENHUSEN, R. SCHAEFER, and T. LANG, "Traffic data from cell phones: a comparison with loops and probe vehicle data," Institute of Transport Research German Aerospace Center, Germany, Tech. Rep., 2003.

[15] J. Kammann, M. Angermann, and B. Lami, "A new mobility model based on maps," *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 5, pp. 3045 – 3049, 2003.

[16] B. White and M. Shah, "Automatically tuning background subtraction parameters using particle swarm optimization," in *IEEE International Conference on Multimedia and Expo*, 2007, pp. 1826–1829.

[17] X. Li, F. Yu, and Y. Wang, "PSO algorithm based online self-tuning of pid controller," in *CIS 2007, International Conference on Computational Intelligence and Security.*, 2007, pp. 128–132.

[18] J. Kennedy and R. C. Eberhart, *Swarm Intelligence*. Morgan Kaufmann, 2001.