DERIVING TRAFFIC DATA FROM A CELLULAR NETWORK

Johan P. Wideberg*, Noelia Caceres, Francisco G. Benitez
Transportation Engineering
ESI- University of Seville
Descubrimientos s/n, 41092 Seville
Spain
*wideberg@esi.us.es

ABSTRACT

Acquiring high quality origin destination information for the vehicle traffic in a geographic area is a tedious and costly task. Traditional methods are expensive, time-consuming and generally only present a snapshot of the traffic situation at a certain time. The technique developed in this paper exploits the use of data already at hand in a GSM network. Instead of monitoring vehicles flows, mobile phones flows are measured and are correlated to the traffic flow. This methodology is based on the fact that a GSM network always knows an estimated position of each terminal, referred to the location area of the base station that provides services to it. For a pilot study a GSM network simulator has been designed to generate a synthetic database with location registers, which is then processed mathematically and transformed into traffic data. Primary results show great potential of this method.

KEYWORDS

OD matrix, traffic monitoring, mobile phone tracking, network simulation, mobile phone positioning.

INTRODUCTION

Origin-Destination matrices are important for designing, analysing and enhancing road networks. These matrices give information about trips for an area of interest, providing an estimate of the number of vehicles travelling between points in this area over a period of time. An OD matrix is difficult and often costly to obtain by traditional methods such as household surveys or roadside monitoring. In addition to the high costs, they are sampling of data is very time-consuming and generally only present a snapshot of the traffic situation at a certain time.

In recent years mobile phone services have become widespread all over the world. In order to provide services to mobile phone subscriber, mobile phone networks have an inherent system to identify the current location of each mobile phone. Many applications will be available in the future exploiting this location information. One of these applications can be used in the field of surface transportation. The novel method proposed in this paper uses information about mobile phones locations to create the traffic data. If a mobile phone is present in a vehicle and switched on, that vehicle becomes a "probe vehicle". Then, mobility information will be obtained analyzing the location of mobile phones on board of them (probe vehicles).

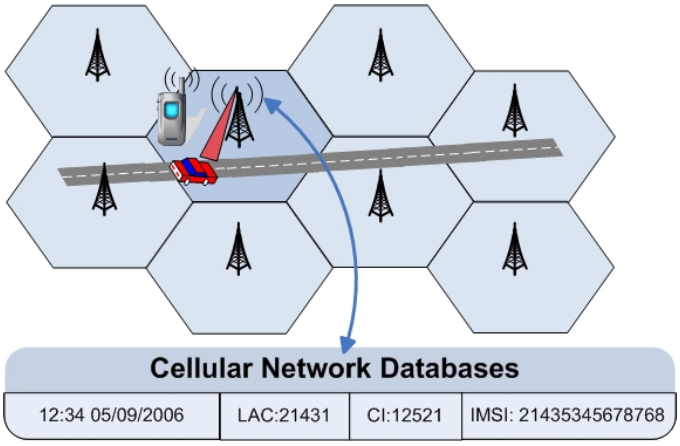

Figure 1 - Vehicle with mobile phone

State of the Art

A fair number of studies relating to this matter have been published in recent years. Bolla [1] presented a model for estimating traffic by means of an algorithm that calculates traffic parameters on the basis of telephone localization data. Lovell carried out various studies [2] [3] on the basis of anonymous data on the position of mobile phones with a view to transport applications such as journey times or speeds. We must also refer to the investigations by Sohn [4] and Akin [5] on O-D matrix calculations using simulations of mobile phone data, as well as other studies focussed on the efficacy of the technique, such as that by Cayford [6], which analyzed the main parameters to be taken into account, namely: precision, metering frequency and the number of localizations necessary to achieve accurate traffic description.

Several pilot projects have been undertaken to study this technique, some of them using the actual localization carried out by GSM operators when a mobile phone is simply switched [7] on and others employing certain data associated with specific situations such as calls (data from phone bills) or handovers [8] [9] [10]. There have even been projects that have developed their own software to capture mobile phone signals [11] [12] [13]. All of them have focussed on the use of localization data associated with mobile phones to obtain information on vehicle movements, while respecting at the privacy of the mobile phone users.

MOBILITY MANAGEMENT IN CELLULAR NETWORKS

One of the main features of mobile phone networks is the fact that mobile stations have no permanent connection to the system. In addition, this system has to deal with transitory users. In order to provide services to its mobile subscribers with short delays and a low signalling cost, the mobile phone network has to track them to be able to locate the base station (cell) in which each phone is currently located. This is why location areas and cells are introduced. In GSM terminology, the coverage area is divided into smaller areas of hexagonal shape, referred to as cells. A cell is defined as the area in which a phone can communicate with a certain base station. A location area (LA) is a geographic area covered by base stations belonging to the same group. The identifier of LA associated to each group is called "Location Area Identifier" (LAI). Other identifier is the "Cell Identity" (CI) that uniquely identifies a cell in a location area. A cell within a GSM network is identified by Cell Global Identifier (CGI: LAI+CI).

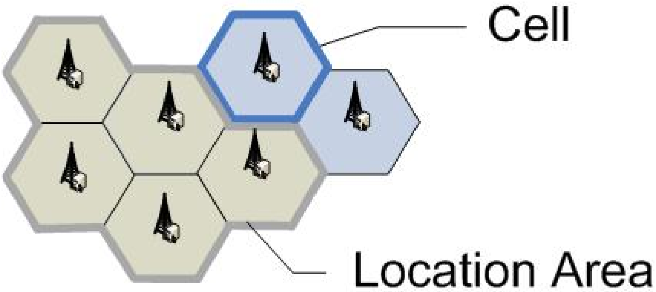

Figure 2 – Cells and location areas definitions

As the subscribers are free to go everywhere in the service area, it is obvious that they will enter and leave cells. Therefore, location information must be managed. The mobility management is the set of processes used to allow the users to be reached wherever they are in the network coverage area. One of the operations involved with mobility management is the location updating. The location updating refers to procedure whereby the phone informs the network about its current location by means of a trigger. The location updating can be divided in two steps. The first consists of sending signalling messages from a mobile station to update its location. The second is the location registration, which consists of a set of processes which updates the network databases, according to location messages which are sent by the mobile stations. This location information is updated and stored in the HLR and the VLRs. The "location updating" procedure is carried out when certain events are produced, for instance a phone moves into a cell which belongs to a new location area or a timer associated with a location process expires. The location information, which is updated in this procedure, is the LAI and CI. These identifiers are corresponded with the location information of the base station (cell) in which the phone has been registered.

New alternative method

As regards the overall topology of the network, its coverage area will be extensive in cities and along the main elements of the transport infrastructure e.g. railway lines, trunk roads and the like. In the case of a road, the presence of mobile phones in vehicles may be used to estimate real time traffic conditions. As mentioned above, the GSM network produces an estimation of a mobile phone position when it is switched on within the network's coverage. These estimated positions refer to the cell and location area of the base station in where each mobile phone has been connected (registered) and they are updated and stored periodically in the network databases (HLR y VLR).

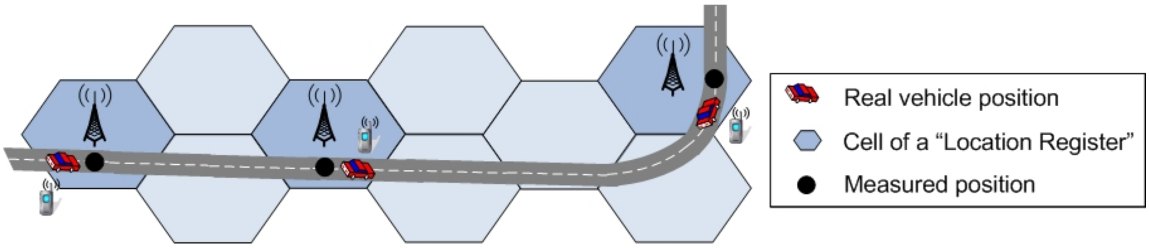

Figure 3 - An example of vehicle trajectory

It is possible to build trips by means of the matched phone positions combining the cells where a mobile phone has been registered and carrying out a map matching process, from which the origin destination matrix can be inferred. It is necessary to bear in mind that this method is only study the trip itself and not the traveller. To keep the personal privacy, the mobile phone identity is encrypted using a unidirectional identifier thus, once encrypted, cannot be recovered.

PILOT STUDY

The aim of the pilot study is to show the feasibility of obtaining traffic counts data and to update the OD matrix from anonymous phones which are on board the vehicles. For a first pilot study, a GSM network simulator has been designed and a synthetic database was generated. The latter is then processed mathematically and turned into an origin–destination matrix and traffic counts.

GSM Network Simulator

Cellular networks have a database that records the current location of mobile phones over time. The phone company has to know the location of all mobile phones at any time to provide a good service to its subscribers. The simulator generates a synthetic database with the location data of mobile phones that are switched on during a day, emulating a GSM network database. The simulator emulates a corridor in an inter-urban area covered by a cellular network. This corridor has a set of cells and location areas which are spread over the road. Each cell covers a road link, along which the simulated vehicle types are travelling.

The simulator creates a set of vehicles with mobile phones on board according to mathematically and statistical criteria. The size of this set of vehicles is estimated using data from real link counts which are situated in the road of interest. The simulator manages the movements of this set of vehicles during a day and events which are related to the mobile phones on board of the vehicles, for instance switching the phone on or off, making or receiving calls, sms.... It also manages the “Location Updating” procedure when a register event is produced, e.g. a mobile station moves into a cell which belongs to a new location area. Each vehicle has an assigned trip to carry out over the corridor, an instant to start its trip and a number of on board mobile phones according to its occupants. So that, mobile phones present in vehicles will be moved within the cells that provide the coverage to the corridor. If it is produced some register event while a phone is within a cell, the mobile phone location will be recorded in the synthetic database to update its location, such as real GSM “Location Updating” procedure. Finally, the database will have stored the location of all phones on board of simulated vehicles during a day. This location data is related to the cell and location area in which a mobile phone has been registered, like a real GSM networks.

Table 1 – Example of “location register” generated by the simulator

Time	*Encrypted ID phone*	*LAI*	*CI*
2006-02-10 07:18:00	11303	2140155102	48751
2006-02-10 07:47:00	94499	2140156021	49833
2006-02-10 08:20:00	97404	2140155901	49301

It is necessary to bear in mind that the phone number or the subscriber name is not used to carry out the analysis. Only it is necessary an identifier to distinguish between all mobile phones. The technical protects the privacy of mobile phone users by encrypting phone numbers as soon as the tracking data is received. So that, all phone numbers in the database are encrypted using a unidirectional identifier to avoid any legal issues.

OD matrix inference

To develop this method, a synthetic database with the location records during a day is used. In this database, the records are already filtered only considers the mobile phones which are travelled through the area of interest. Analysing the synthetic data generated by the simulator, the mobile phones that have been moved into the cells that cover the interested road is recovered. Combining the cells where each mobile station has been registered, it is possible to build a trip, from which the location area origin and destination may be inferred, which started at the moment when the first register was recorded and finished with the last register.

Traffic Counts

For the purposes of determining traffic flow, the process is based fundamentally on the fact that mobile terminals perform the whole “location updating” procedure (registration) when they enter a location area with an identifier distinct from the one they had previously recorded. For this reason, irrespective of whether the terminal is recorded in the network due to any other event, the records for a given phone will always be available when it enters a new location area, as long as it is switched on, whether or not it is being used.

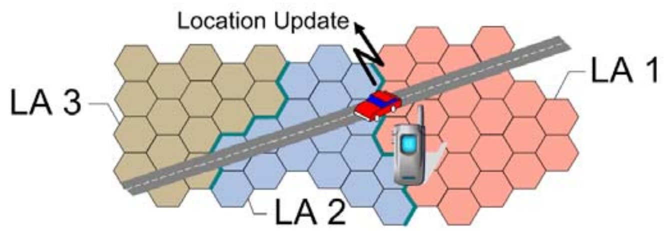

Figure 4 - Example of Location Updating on entering a new LA

Estimations of traffic volumes passing through these borders are carried out by looking for the moments in which registrations take place due to changes of LA, that is to say moments in time when the first registration occurs with a LAI different from that previously recorded. In this way passage between areas is monitored as if traffic counts were involved.

Adjustment Factor (Mobile phones per Vehicle Equivalents)

It is important to bear in mind that this technique generates data associated with mobile phones switched on for a particular network provider. However, the main interest is getting vehicle traffic data, not mobile phones data. Consequently, it is used an adjustment factor that relates vehicles to mobile phones switched on. This factor (f_{MPV}) is used in the GSM data analysis to convert phones data into equivalent vehicles data. It is defined as equation (1):

$$CellPhones_{ON\ Op} = CellPhones_{ON\ Op}\Big|_{vehicle} \cdot Vehicles \Rightarrow f_{MPV} = \frac{1}{CellPhones_{ON\ Op}\Big|_{vehicle}} \quad (1)$$

The concept of estimating mobile phones per vehicle equivalent (MPVE) is to calculate approximately the number of mobile phones, which are switched on and belong to a specific network provider, displaced on board of each vehicle under a specified area. This value of mobile phones per vehicle is defined using statistical parameters of a mobile phone market from annual studies published by the Telecommunications Market Commission in Spain (Mobile Phone Penetration, Operator Market Share by Number of Subscribers, ...), together with other traffic vehicular parameters (vehicle occupancy, proportion of vehicle types, etc).

RESULTS

The aim of this report is to show the feasibility of using the database of mobile phones locations to infer OD matrix by means of phones. This data could be extracted directly from a mobile phone network but to do that cooperation with a phone company would be necessary. Before entering any agreements we have opted to do a pilot study using a synthetic database generated by a simulator.

The results obtained from the analysis of this database provide an estimate of the number of phones travelling between points on a network over a given period of time. The phones trips are then turned into vehicles trips by an “adjustment factor”. These estimated vehicles trips can be validated by means of the real ones generated by GSM Network simulator, since the simulator also stored the real trips in order to make the comparison with the estimated ones.

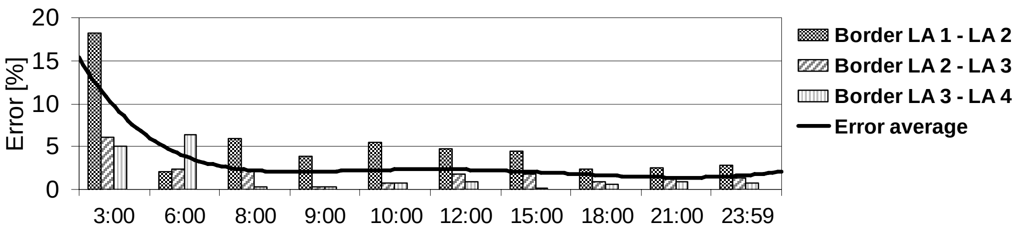

Figure 5 - Traffic Count Estimation Error from simulation over each period

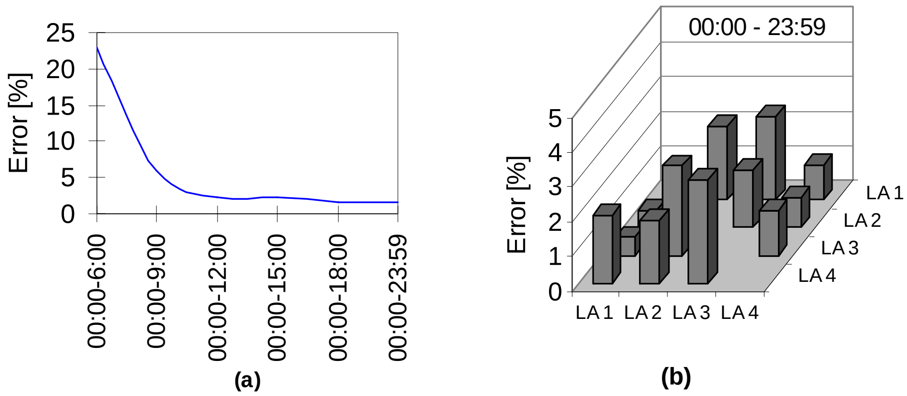

Figure 6 - O-D Matrix Estimation Error from simulation over different periods.

This error is originated principally by the "adjustment factor" used to transform phones trips to vehicles trips, which is calculated using average statistical figures evaluated in one day. It can be seen that the correction adjustment is almost perfect, producing significantly reduced levels of error. Logically, this depends on the factor employed, which is calculated using the same average values as those used to define the statistical distributions modelling the simulator's vehicles and mobile phones. Obviously, calculation of the adjustment factor will not be so trivial when real data are used, enabling more complex analysis to be undertaken.

Other factors also influence in this estimation error. One of them is due to mobile phones that are switched off during a whole. These phones provide erroneous information because that trip is not detected. Moreover, if a phone is switched on during a trip, the estimated trip is not either the real one because it is considered that started when the phone switched on but not when the vehicle began the movement.

However, the reduced error levels obtained in multiple simulations and analysis have demonstrated that it is possible to obtain traffic data from a cellular network, show the potential of this technique to collect traffic data of vehicles using the location data of phones on board of them.

CONCLUSIONS

The use of mobile phones on board of vehicles such as measuring tools (without revealing confidential data or affecting its performance) presents various advantages with regard to traditional methods. The main one lies in the possibility of estimating traffic data with low deployment cost, avoiding the need for installing additional devices along the road network or the use of other costly methods such as traffic surveys. The extracted data come from the same terminals that most people use nowadays, by the simple fact that the phones are switched on in any zone with GSM coverage. Consequently, and bearing in mind the fact that the process is fast and supposes to handle a widespread sample of people, this technique based on anonymous mobile phone data represents a major revolution in traffic mobility studies.

REFERENCES

[1] Bolla R., Davoli F., Giordano A. (2000). "Estimating road traffic parameters from mobile communications". In Proceedings *7th World Congress on ITS*, Turin, Italy.

[2] Lovell David J. (2001). "Accuracy of speed measurements from cellular phone vehicle location systems". *ITS Journal* Vol. 6, No. 4, pp. 303-325.

[3] Smith Brian L., Pack Michael L., Lovell David J. and Sermons M. William (2001). "Transportation management applications of anonymous mobile call sampling". In Proceedings *11th Annual Meeting of ITS* America, Miami, FL.

[4] Sohn K. (2004). Dynamic Estimation of Origin-Destination flows Using Mobile phones as Probes. SDI 2004-R-04. Department of Urban Transportation. Seoul Development Institute. Korea.

[5] Akin D. and Sisiopiku V.P. (2002). "Estimating Origin-Destination Matrices Using Location Information from Mobile phones". *Proceedings of 49th Annual North American Meetings of the Regional Science Association International*, Puerto Rico.

[6] Cayford R. and Johnson T. (2003). "Operational Parameters Affecting Use of Anonymous Mobile phone Tracking for Generating Traffic Information". *82th TRB Annual Meeting*.

[7] ITIS Holdings - Press Releases. "Estimotion Technology chosen for traffic collection in Antwerp". Available: http://www.itisholdings.com/pdf/itisproximus.pdf

[8] "Mobile phones generate traffic information in Brabant". Mobile Traffic Service (MTS), development by LogicaCMG - http://www.mts-live.com

[9] Ratti C., Pulselli R. M., Williams S., Frenchman D. (2005). "Mobile Landscapes: using location data from mobile phones for urban analysis". *Environment and Planning B – Planning and Design*.

[10] Thiessenhusen, K-U., Schäfer R-P. and Lang T. (2003). "Traffic Data from Mobile phones: A Comparison with Loops and Probe Vehicle Data". Institute of Transport Research, German Aerospace Center, Germany.

[11] Airsage - Press Releases. "Using Cellular Technology to Provide an Intelligent Traffic Roadmap". Available: www.airsage.com/assets/AirSageExecutiveSummary.pdf

[12] Airsage - Press Releases. "Mobile phones could keep you out of traffic jams". Available: www.airsage.com/assets/Virginian_Pilot_4_27_05.pdf

[13] Delcan Corporation – Press releases. "Traveler Information Derived from Mobile phones Begins to flow from Innovative Public-Private Partnership in Baltimore". Available: http://www.delcan.com/prod/index.php?id=295