TRAFFIC VOLUME ESTIMATION THROUGH CELLULAR NETWORKS HANDOVER INFORMATION

Merkebe Getachew Demissie, Civil Engineering Department, Coimbra University, Portugal, Email: demissie@dei.uc.pt

Gonçalo Homem de Almeida Correia, Civil Engineering Department, Coimbra University, Portugal, Email: gcorreia@dec.uc.pt

Carlos Lisboa Bento, Informatics Engineering Department, Coimbra University, Portugal, Email: bento@dei.uc.pt

ABSTRACT

Traffic management agencies carry out different types of traffic counts. Permanent traffic counter (PTC) is the preferred counter that provides traffic statistics throughout the year. Due to its expensive installation and maintenance costs, PTC has limited road network coverage; therefore, agencies choose to utilize sample traffic counts from seasonal traffic counter (STC). However, road network covered with STC lack continuous observation to update changes in traffic in a regular basis, which causes a series impediment to the effective use of active traffic management schemes. One way that has been tried to obtain complementary traffic information is the use of cellular networks. This paper presents a method for estimating hourly traffic volume through a combined use of average city-wide traffic volume and cellular networks handover information. To test this method, handover data were collected from cellular towers in Lisbon, Portugal, which are situated along the Eixo norte-sul itinerarios principais road. The traffic volumes were also obtained from traffic counters installed in the same road. First, we applied statistical analysis to investigate the relationship between vehicular and cellular traffic. Then, we built a multiple linear regression model. Results from statistical analysis proved that handover can be used as a proxy to detect vehicle movement. We used mean absolute percentage error (MAPE) for model evaluation. The MAPE of the regression model is 18% less than the MAPE obtained through the estimation method that only based on average city-wide traffic volumes. This seems to indicate that the feasibility of the combined use of handover and traffic volume providing better accuracy in capturing site-specific traffic profile compared with the typical average city-wide traffic volume. It can be concluded that this study encourages the exploration of the use of cellular network handover information in estimating road traffic volume.

Keywords: handover, cellphone, city dynamics, cellular networks, traffic estimation

INTRODUCTION

The steadily increasing traffic congestion associated with the rapid urbanization has caused slower speed, queue formation, and increased travel times, which impose costs on the economy and generate multiple impacts on urban environment. Congestion can be reduced through employment of operational traffic management measures that can optimize the use of existing infrastructure to insure high levels of service, however, this needs acquiring of reliable and detailed traffic information. Traffic management sectors use several techniques to gather raw traffic information that basically fall into two major groups: point-detection and vehicle-based detection systems (OECD, 2007). A single set of traffic data collection system has functional limitations; that is, it does not offer the amount of data required to have a realistic and comprehensive view of the traffic stream in urban areas.

The interactions within the traffic system can be quantified by two groups of traffic stream parameters. Volume or flow rate, speed, and density are the macroscopic parameters that describe aggregate traffic flow characteristics and relationships. On the other hand, headway, gap, and lane occupancy are the microscopic parameters that describe the characteristics of individual vehicles (Abdulhai and Kattan, 2004). Understanding what type of flow is occurring in a given situation helps to decide the most relevant analysis methods. A principal road in an urban environment is used to conduct the present study, and traffic volume is the preferred parameter to characterize the traffic system.

Traffic volume and flow rate are measures that quantify the amount of traffic traversing a point in a roadway system per unit of time. Traffic volume studies are conducted to determine the number, movements, and classifications of roadway vehicles at a given location. These data can help identify critical flow time periods, determine the influence of large vehicles or pedestrians on vehicular traffic flow. The duration of the traffic count defines the type of traffic volume: annual, daily, hourly, etc. Flow rate represents the equivalent hourly rate of vehicles traversing a roadway system during a time interval of less than 1 hour (TRB, 2000). Annual average daily traffic and average daily vehicle distance traveled are the two most used traffic statistics mainly for traffic planning purposes. However, traffic managing authorities should also collect hourly traffic volumes to look up to the operational characteristics of the road at different times of a day (FHWA, 2001).

The current practice for data collection regarding traffic volume depends on either human observation or different forms of remote sensing (e.g. inductive loops, automatic video feed-based counts, etc.). The use of on-road sensors (e.g. inductive loops) for collecting traffic data is necessary but not sufficient because of their expensive implementation and maintenance cost that caused limited road coverage. In the absence of a traffic count from a specific site, prediction is usually made through the average traffic obtained from other places or historical traffic data of a given location. However, this forces traffic management authorities to rely on an incomplete picture of the traffic stream in the city (OECD, 2007).

The development of intelligent transportation systems requires heterogeneous and high quality traffic information. This has put enough pressure to the road managers to seek for additional traffic information sources (Leduc, 2008). One way that has been tried to obtain this information is through the use of cellular networks. In comparison to on-road sensors, cellular networks provide mobility related events that can be obtained during conventional operation, such as Location Area (LA) update, Route Area (RA) update, and cell update (handover) (Valerio et al., 2009).

In this paper we explore a hybrid use of cellular networks handover information and traffic volumes from inductive loops for the purpose of traffic volume estimation. Handover is a cell based location update, which is the process of transferring an on-going call or data session from one area to another without interruption of service (Zeng and Agrawal, 2002). Handover is measured regularly to supervise the location management schemes in the cellular network. Though handover information is extracted without incurring additional costs and operational risks for the network, it has a limitation of recording locations at the granularity of a cell tower location, which gives uncertainty on the exact location of the handover events. However, the aggregate handover data obtained for this study remains attractive for our analysis at the scale where this detailed level of location precision is unnecessary.

The remainder of this paper is organized as follows: Section 2 presents review of related works. Section 3 describes the case study area and dataset collection procedures. Section 4 and 5 present the handover based system and methods followed in this study. Section 6 presents the result and discussion. Finally, the paper ends with the main conclusions that it was possible to withdraw from the study and presents future research directions.

LITERATURE REVIEW

Traffic estimation has been an important issue along the last decades. Greenshields conducted one of the first empirical studies on measurement of traffic volume, traffic density and speed through a snapshot of traffic by an aerial camera in the 1930s (Kühne, 2008). The development of intelligent transportation systems introduced new ways of obtaining road traffic data from alternative sources. The study by Varaiya et al. (2008) used wireless magnetic sensor networks to detect the presence and movement of vehicles in real time. Herrera et al. (2010) and Ahas et al. (2010) used GPS-equipped phones as probes to gather mobility related information within a cellular network.

In recent decades, several studies regarding the analysis of data obtained from cell phone use have been carried out for different purposes. A number of Studies (Alger et al., 2005; Bar-Gera, 2007; Liu et al., 2008) primarily used handover data to estimate speed and travel time. In Alger et al. (2005), the estimated speed by the data had revealed a bimodal pattern with a double peak originated on lorries at about 80-85 km/hr and on cars at about 120 km/hr. The comparison between speed and travel time estimated from the cellular network and loop detector data sources was carried out by Bar-Gera (2007). This study considered 20,368 common time intervals from both sources and achieved an average absolute relative difference of 10.7%. A similar study by Liu et al.(2008) compared speed estimate of the cellular network against speed measured by loop detector and provided results in terms of variation of speeds at different levels of congestion. The speed difference between the two methods was around 10 mph (16.1 km/hr) on free flowing traffic state. However, during light and moderate congestion, where traffic conditions generally fluctuate the most, speed difference of more than 20 mph (32 km/hr) was registered. Herrera et al. (2010) deployed 100 vehicles carrying a GPS-enabled Nokia N95 phone to acquire continuous location and speed profile of the vehicles. This data demonstrated the feasibility of the proposed system for real-time traffic monitoring. Results suggested that 2 to 3% penetrations of cell phones are required to provide accurate measurements of the velocity of the traffic stream. Handover is used for speed and travel time estimation with the advantage of obtaining data from a wide transport network, however, the limited geographical resolution that puts multiple routes within a cell causes error for speed and travel time estimation purposes (Caceres et al., 2008).

The study by Thajchayapong et al. (2006) detected traffic congestion through a correlation between extreme cell dwell times and an incident in a road traffic stream. A similar study by Hongsakham et al. (2008) applied K-means clustering and backpropagation neural networks to estimate the degree of road traffic congestion through the use of cell dwell time. The results showed that the classification accuracy by K-means clustering and neural networks were 59.14% and 88.65% respectively.

A study by Caceres et al. (2007) used a Global System for Mobile communications (GSM) micro-simulator to estimate traffic volume through LA update information of the cellular network. Good estimation accuracy was achieved by comparing the phone movement, which was translated to vehicle movement using a suitable adjustment factor against the vehicle movement stored in the GSM simulator. The study by Thiessenhusen et al. (2003) discovered the coexistence of handover and traffic volume peaks during the morning and afternoon rush hours.

Different types of cellular network information were applied for estimating origin-destination (OD) matrices. In Pan et al.(2006), a cellular based approach was used to extract mobile phone positions every two hours to infer trip distribution data. White and Wells (2002) obtained cellular data that was stored for billing purpose, which consisted of phone position at the start and end of the call and call duration, for computing OD flows. The usual cellular network information that can be used to infer OD matrices are handover and other location updates with the advantage of obtaining good sample size and data collected directly from the traffic stream. However, it is not possible to detect the intra-area trips in the case of a single cell covering a large area (Caceres et al., 2008).

Regardless of the important efforts in applying cellular networks information for the development of speed, density, OD flows, and travel time estimation, by far cellular networks has been the less exploited source of data for the purpose of traffic volume estimation and there is still work to be done. In this work, an innovative method of traffic volume estimation method is developed. We explore the combined use of traffic data from traditional on-road sensors and cellular networks to provide high quality traffic information that can be utilized by transportation sectors.

DATA COLLECTION

The study site is identified inside the Lisbon metropolitan area, which comprises of the municipality of Lisbon and its surrounding. Lisbon is the capital of Portugal and the centre of the Lisbon metropolitan area. Lisbon is chosen because of its high level of cell phone use and its city dynamics. According to statistics from ANACOM (2010), active mobile telephone cards per 100 Portuguese inhabitants grew to 159.9 by the end of year 2010, from 140.4 in the year 2008. In this study handover data was obtained from TMN Company. In December 31, 2010, TMN had 7.42 million subscribers in Portugal, which accounts for 45% of the total mobile subscribers in Portugal. TMN uses GSM and Universal Mobile Telecommunications System (UMTS) technologies to provide mobile communication services.

The first set of data, handover information, was extracted from 487 cell towers that carry approximately 1669 cell sectors pointed in various directions in Lisbon. The hourly handover counts were gathered on the April 5, 6 and 12 of 2010. The handover data was the one generated through voice traffic. The handover map (Figure 1) shows the flow of calls from one cellular tower to the other. The triangles are representing the location of the cell towers, and the trajectories are showing the availability of handover connection.

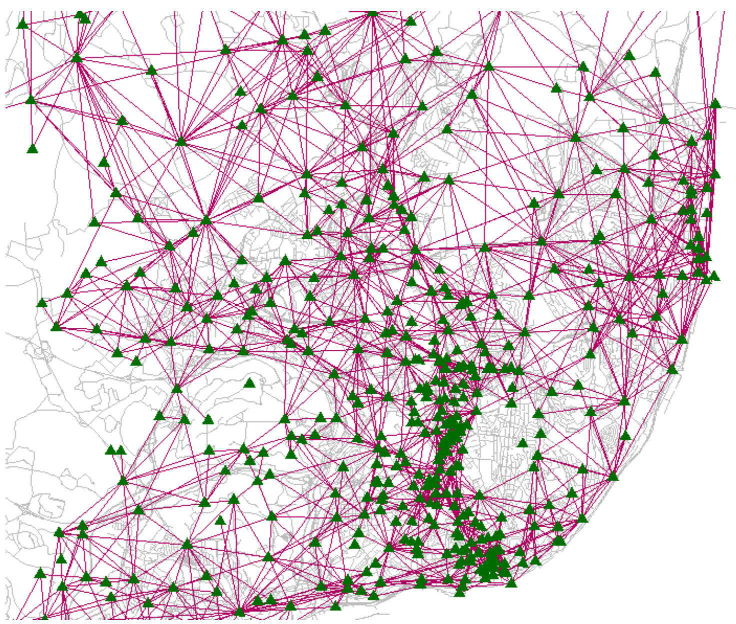


Figure 1 - Hanover map

The second set of data, the traffic volume, was also obtained from week days in same days as the handover data, April 5 and April 6 of 2010. In April 5 of 2010 hourly traffic volumes were obtained from 101 traffic counters equipped with inductive loops that were made available to us by the municipality of Lisbon. Out of the 101 traffic counters, traffic data from the 94 counters were used to prepare the average city-wide traffic volume, and the remaining traffic volumes from the seven counters were used to build a regression model that estimates traffic volumes. In April 6 of 2010 we obtained hourly traffic volumes from the same 7 counters that were used to build the regression model to use them as model validation dataset.

Figure 2 shows geocoded points of the traffic counters available in the city that are represented by circles and cellular tower locations by triangles. In addition, the figure 2 exhibits a detailed representation of the traffic counter locations (circles) on the Eixo norte-sul itinerarios principais road, where the road is designated by blue color. The direction of the traffic movement is indicated with an arrow next to the counters considered in the study for model development and validation. The Eixo norte-sul itinerarios principais road is set to be of restricted access that forbids movement of pedestrian, animal and bicycle traffic. The study segment has a length of 16km in one direction; it has an annual average daily traffic of 76731 vehicles computed from all the counters along this road in the year 2010.

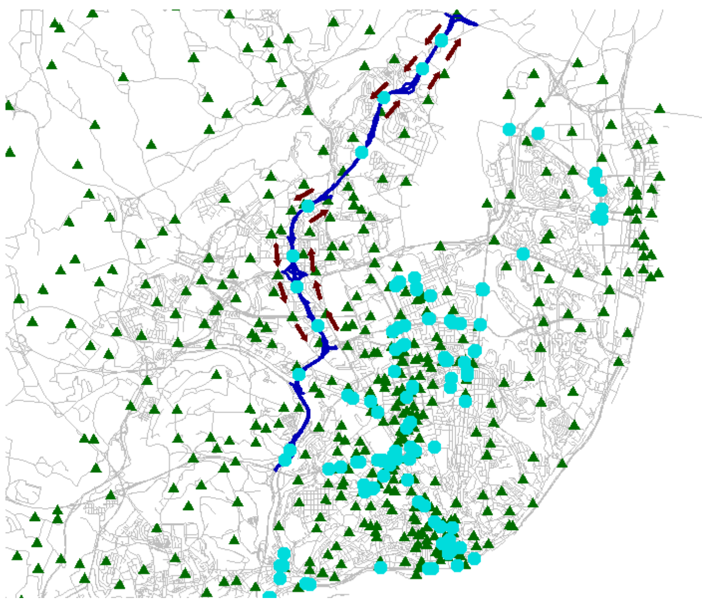


Figure 2 - Detailed representations of traffic counter locations on the Eixo norte-sul itinerarios principais road

DESCRIPTION OF THE HANDOVER BASED SYSTEM

The advancement of transportation through the application of innovative technologies that enable various users to make better use of transportation networks has induced more trips. In order to accommodate these mobility features of travelers a cellular network carries out a series of operations, such as location update (LU) or handover. A cellular communication system has to track the location of its user in order to forward calls to a relevant cell within a network and location management is a scheme responsible to carry out such operation. Cells within a network are grouped to form a mega cell, which is called location area (LA) (Figure

3). Therefore, a mobile phone within the boundary of a given LA can move without updating its location until transitioning to a new LA.

LU is a process used to inform the network of a mobile phone´s location. This process requires the phone to register its new location with the current base station (cell), such that the forwarding of incoming calls would be performed. The status of the mobile phone determines the rate of LU. During an ongoing call the network always knows the exact cell where the phone is connected through the handover process. On the other hand, when the mobile phone is in its idle status (turned on, but not on-call) the network do not have the location of the exact cell; instead the network knows the current LA through LU process. If a call is to be forwarded to a user, the network must page every cell within the LA to determine their precise location (cell).

The handover process is fired by the operational rules configured in the cellular networks. For simplification we represented the cell as hexagonal area (Figure 3), however, the dimension of a cell depends on the geography of the area and cellular use within a cell Zeng and Agrawal (2002).

The handover process is explained with an example of a trip maker who is on active call and driving in the eastbound direction between cell 1 and 6. As the vehicle crosses the roadway from cell 1 to cell 2, the call must be transferred between cells without interruption. Therefore, a handover event takes place handover 1(HO1). This event occurs afterward at the location of HO2 to HO6 as far as the call is in progress. The idea behind the handover-based system is to use the handover lines at the cell boundaries as “virtual” traffic counters. These virtual traffic counters are assumed to capture the vehicle movement instead of using traditional on-road sensors represented by the circles (T1 to T5).

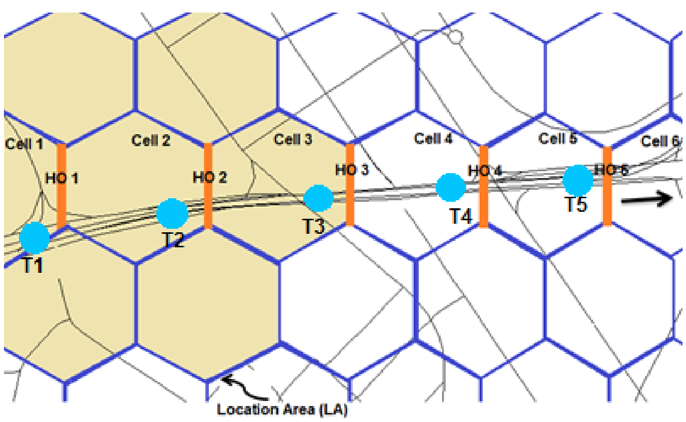


Figure 3 - Schematic representation of handover-based system

METHODS

Visualization

GIS software is selected from a pool of specialized tools that has been prominent for this kind of analysis. Several procedures were developed using the best functionalities of GIS software (Geomedia Professional) through its urban transportation analysis package. The road network of Lisbon is used as the underlying digital map in this study. The digital road map contains all the different levels of roads from high to low mobility. The primary target of this analysis is to investigate the use of handover data to infer information about vehicular traffic in the vicinity of the antennas. In addition, we explored the use of cellular networks handover information for traffic volume estimation in urban environment with the objective of assessing its possibilities and pointing directions for future research in this field. In order to verify our assumption, different geographic representations were produced through the GIS that allowed us to visualize various characteristics of call movements. Using GIS, handover events from the entire cell sectors sited in the same location were summed up and grouped into incoming and outgoing handovers for the same location.

Statistical analysis

Besides the visualization task, statistical analysis was performed that allows us to verify a relationship quantitatively. This study raised the following question:

- Is the location of cellular towers accommodating high number of moving calls significantly closer to the main road links when compared to the location of towers serving fewer movements?

In order to answer this question, a statistical method based on the comparison of two sample means was established. Two mutually exclusive hypotheses are required to examine the difference between two means: null hypothesis (H_0) that assumes that both samples have equal means and an alternative hypothesis (H_1) that assumes that both means are different. The two population means, μ_1 and μ_2 are obtained from the sample of a given population designed in each case. The two-sample t-test is applied to determine if the differences between the means in the samples are statistically significant. The sampled data were analyzed by first computing the standard error (SE), degrees of freedom (DF), test statistic, and the P-value associated with the test statistic($Eqn. 1\ to\ 3$).

The standard error of the sampling distribution is computed as:

$$\mathrm{SE} = \sqrt{[(\mathrm{s}_1^2/\mathrm{n}_1) + (\mathrm{s}_2^2/\mathrm{n}_2)]} \qquad \text{Eqn. 1}$$

Where, s_1 is the standard deviation of sample 1, s_2 is the standard deviation of sample 2, n_1 is the size of sample 1, and n_2 is the size of sample 2.

The number of degree of freedom is computed as:

$$\mathrm{DF} = \left(\frac{\mathrm{s}_1^2}{\mathrm{n}_1} + \frac{\mathrm{s}_2^2}{\mathrm{n}_2}\right)^2 / \left\{\left[\left(\frac{\mathrm{s}_1^2}{\mathrm{n}_1}\right)^2 / (\mathrm{n}_1 - 1)\right] + \left[\left(\frac{\mathrm{s}_2^2}{\mathrm{n}_2}\right)^2 / (\mathrm{n}_2 - 1)\right]\right\} \qquad \text{Eqn. 2}$$

Where, s_1 is the standard deviation of sample 1, s_2 is the standard deviation of sample 2, n_1 is the size of sample 1, and n_2 is the size of sample 2.
The test statistics, which is a t-score (t), is computed as:

$$t = [(\mu_1 - \mu_2) - d]/SE \qquad \text{Eqn. 3}$$

Where, μ_1 is the mean of sample 1, μ_2 is the mean of sample 2, d is the hypothesized difference between population means, and SE is the standard error.

Regression model

Regression analysis is a statistical technique for estimating the relationships between one variable and another set of variables. In a simple linear regression model, a single response measurement y is related to a single predictor x for each observation: $y = \alpha + \beta x$. In most problems, more than one predictor variable will be available. This leads to multiple regression: $y = \alpha + \beta_1 x_1 + \cdots + \beta_n x_n$. Where, α is the intercept and β_1 to β_n are unknown parameters to be estimated.

RESULTS AND DISCUSSION

Visualization

Three different patterns of cell phone movements are revealed (Figure 4), which are measured by the number of active calls that move from one cell to another. The yellow and red circles show the total number of outgoing and incoming handovers respectively at a given tower location (all sectors in the tower). Their size is proportional to the percentage that the outgoing and incoming handovers represent of the total sum of handovers so large circles mean that one type of handover direction dominates the other. The orange circles show how balanced the incoming and outgoing handovers are at a given tower location because they result from the overlap of both colors. The main road links of Lisbon (links that have more than 1,500 Veh/hour) is also represented at the background to provide the sense of how handovers can be related to mobility in the city.

The underlying assumption is that the pattern and amount of cell phone traffic movement is related with the intensity of urban movements and understanding this relationship will help in managing urban dynamics. In its most simplistic form, neighborhoods with cellular towers showing a high amount of handover events are likely to have high mobility. Neighborhoods with cellular towers showing a high and balanced number of handovers (equivalent number of incoming and outgoing handovers) are likely to have high mobility passing near them. Neighborhoods with cellular towers showing a high number of incoming handovers are likely to have a high unidirectional mobility pattern towards the neighborhood.

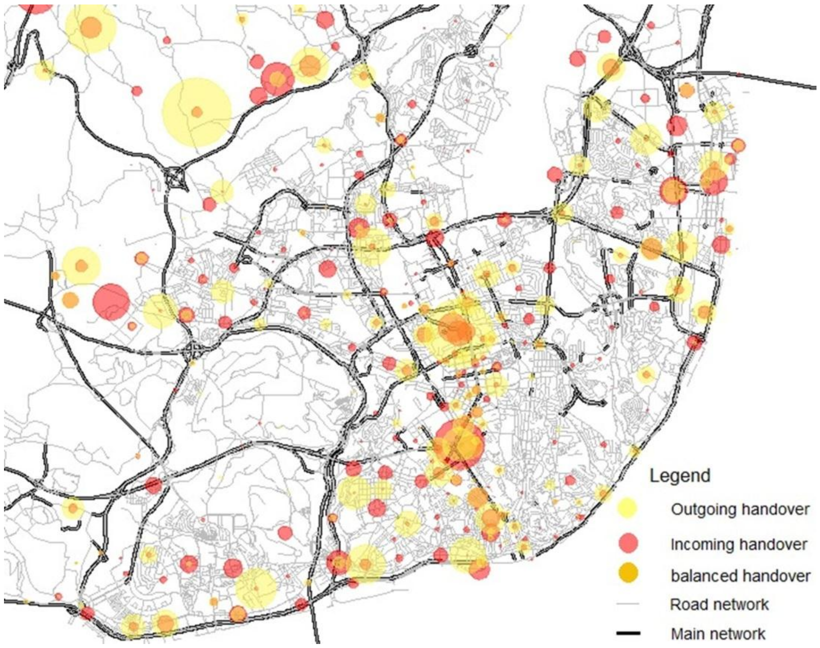


Figure 4 - Incoming and outgoing handovers over Lisbon´s main road links between 8 to 9AM on the 12th of April 2010.

The visualization (Figure 4) gives mainly a qualitative view about the amount and patterns of call movements. Previous studies, using call volume (Pulselli et al., 2008), handover (Ratti et al., 2005), and erlang (Calabrese et al., 2011), also achieved notable visualization results in the area of urban analysis. The standard visualizations obtained are graphically appealing; nonetheless, the relationship between intensity of cellular activities and the characteristics of a given district requires thorough validation.

In the following section, we carry out statistical analysis to investigate an assumption which we formulated, and we perform a correlation analysis to show the association of traffic volume and handover counts. Then a linear regression model was developed aiming at estimating traffic volume in an urban environment through a combined use of handover and average traffic counts.

Statistical analysis

Assumption: It is claimed that cellular towers with a high number of handover events, both incoming and outgoing, should be adjacent to the main road links when compared to the rest of the cellular towers because the first should denote more movement.

The criteria to choose cell towers with equivalent (balanced) high number of incoming and outgoing handovers were the following: (1) the difference between the number of outgoing and incoming handovers has to be less than 30% of the total handover, and (2) the sum of the incoming and outgoing number of handovers has to be more than 144 per hour. The choice of 30% is arbitrary. Out of a total of 29,546 road links with different traffic capacities, road links that with capacities beyond the 75th percentile were considered as main road links and the computed value was 1,500 Veh/hour. The same cut-off point of 75th percentile was applied to the total handover values and it was aligned to the value 144 handovers/hour.

μ_1: Mean distance of towers with high and equivalent incoming and outgoing handovers to the main road links (the distance measures how far apart the locations of the towers and the main road links are).
μ_2: Mean distance of all the other towers to the main road links.

Thus the hypothesis testing is the following:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 < \mu_2$$

Table 1 shows the results of the hypothesis test, where the null hypothesis is rejected with a significance level of 5% in favor of the alternative hypothesis. This confirms the assumption that the average distance of cellular towers accommodating high number of moving calls are closer to the main road links compared to the towers serving fewer movements.

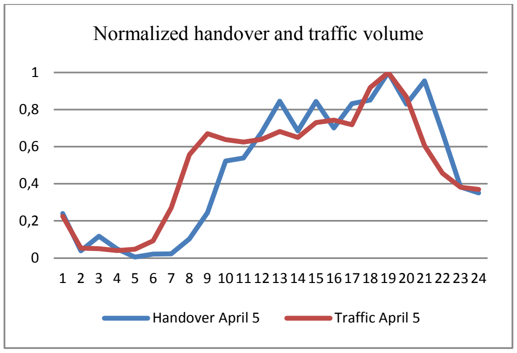
Table 1 - Hypothesis test: distance of balanced towers to the main road links

Groups	Mean (m)	s	n	P-value
Towers with balanced handover	251	247	27	0.0440
All the other towers	341	384	403	

s: standard deviation; n: sample size

Correlation analysis

Figure 5 shows normalized average traffic volumes and handover counts plotted over the hours of two different days, April 5 and 6 of 2010. We investigated the correlation between cellphone handovers and traffic volumes in order to understand the relationship between cellular and vehicular traffic. We compared the hourly handover counts with the corresponding hourly traffic volume counts collected from the 7 traffic counters that will be used for model building and validation. We obtained an average coefficient of correlation of 0.72 and 0.76 respectively for April 5 and April 6 of 2010.



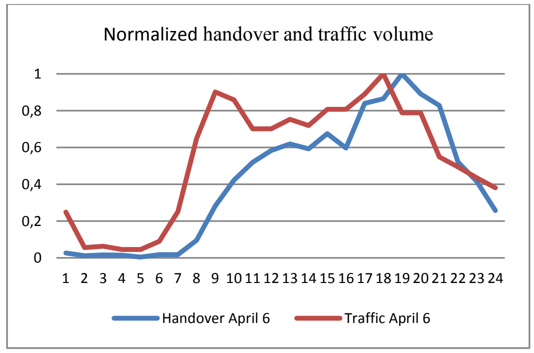


Figure 5 Normalized average traffic volumes and handover counts plotted over the hours of the day, where 1 implies midnight to 1AM, 2 implies 1AM-2AM, and so on, and left is April 5 2010 and right is April 6 2010.

Regression model application

The linear regression model was built using the SPSS software (IBM SPSS Statistics 20). Out of 336 hourly traffic volumes collected from the Eixo norte-sul itinerarios principais road (2 days x 7 counters x 24 hours); the data from April 5, 2010 (168 traffic volumes) were used as dependent variables to build the regression model and half of the remaining data from April 6, 2010 were used for model validation. Two explanatory variables were used to characterize the traffic volume. The first exploratory variable was the average city-wide traffic volume (Avg.T), which was built by averaging traffic volumes obtained from 94 traffic counters in Lisbon during April 5, 2010 (that excludes the traffic counts from the 7 counters used for model validation). The second exploratory variable was the corresponding handover count (HO) from April 5, 2010. Table 2 shows the components of the developed model.

Table 2 - multiple regression model components

Variables	Coefficients	t Stat	P-value
Intercept	54.938	0.550	0.583
HO	20.628	11.404	<0.001
Avg.T	1.061	11.557	<0.001
R Square	0.798		

We used the R-square statistic to evaluate the overall model fit to the data. An R-square value of 0.798 was obtained. Analysis of the estimated coefficients was carried out to understand the relative importance of the explanatory variables that were used to estimate the traffic volume. The p-value for the intercept is greater than 0.05, which shows the coefficient is not different from zero. The coefficients for the HO and Avg.T were both statistically significant at 5% significance level.

The regression model was validated using the handover and traffic volume data from April 6, 2010. The mean absolute percentage error ($MAPE$) has been chosen for model evaluation, which is shown in $Eqn. 4$:

$$MAPE = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{\hat{x}(k) - x(k)}{x(k)}\right| \times 100\% \qquad Eqn. 4$$

Where,

$\hat{x}$ estimated traffic volume

x actual traffic volume

n total number of traffic volume

k Kth traffic volume

We computed the $MAPE$ value for three different time windows. $MAPE$ for the morning peak hours (7AM to 10AM) is 25.6%; $MAPE$ for the afternoon peak hours (5PM to 8PM) is 27.5% and $MAPE$ computed for all the time periods is 28.8%.

In the absence of a traffic count from a specific site, prediction is usually made through the average traffic from other places or historical traffic data of a given location and we call this information average city-wide traffic volume. Therefore, we made a prediction of the traffic volume for April 6, 2010 for the same road (Eixo norte-sul itinerarios principais) based on the average city-wide traffic volume. For our analysis the average city-wide traffic volume was built by averaging traffic volumes obtained from 94 traffic counters in Lisbon during April 5, 2010 (that excludes the traffic counts from the 7 counters used for model validation). We did the validation of this prediction through the same dataset used to validate the regression model. We also computed $MAPE$ for the three time windows for prediction evaluation. $MAPE$ for the morning peak hours (7AM to 10AM) is 39.9%, $MAPE$ for the afternoon peak hours (5PM to 8PM) is 50.9% and $MAPE$ for all the time period is 46.8%.

The $MAPE$ computed for all the prediction times by the regression model is 18% less than the $MAPE$ delivered by the average city-wide traffic volume. This seems to indicate that relying on the average city-wide traffic volume in the absence of traffic data from a specific site would give less accurate prediction. This suggests that the combination of average traffic with the handover count can provide better predictions and the added value of handover data in capturing site-specific traffic profile is indispensable.

CONCLUSIONS AND FUTURE WORKS

The capabilities of conventional road traffic data collection methods (loop detectors, automatic video counts, etc.) are limited due to high installation and maintenance costs and their poor road network coverage. One way to solve this problem can come from using other sources of data. In this study, we explored the combined use of data from loop detector and cellular networks for traffic volume estimation in urban environment with the objective of assessing its possibilities and pointing directions for future research in this field.

Two different analyses were carried out to uncover the mobility relevance of handover information to complement the effort on traffic volume estimation. In the first part of our work we performed two experiments: (1) GIS software was used to manage, analyze and visualize the cell phone data, (2) Statistical analysis and correlation analysis were performed to investigate the relationship between vehicular and cellular traffic. It was found that cellular towers characterized by high and balanced number of incoming and outgoing handovers are located in the vicinity of the main road network thus where there are the major flows of people moving in the city.

In the second part of our work we developed a tentative method aiming at predicting the traffic volume in an urban environment through combination of data from sources: traffic counter and cellular network. For this purpose we built the linear regression model. The performance of our model was validated through an external dataset. We used MAPE for model evaluation. The MAPE obtained through the regression model is 18% less than the MAPE delivered by the average city-wide traffic volume. This seems to indicate that relying on the average city-wide traffic volume in the absence of traffic data from a specific site would give less accurate prediction. This suggests that the combination of average traffic with the handover count can provide better predictions and the added value of handover data in capturing site-specific traffic profile is indispensable.

Our approach is capturing, but it has also the following shortcomings. The handover event is only achieved through phones that are actively making a phone call, and the duration of the associated call must be long enough to traverse the boundaries of two cells. Thus, it was not possible to make a direct correspondence of the cellphone activity (handover counts) and traffic. Another limitation is related to the handover counts. Even using rules to try to associate specific handover events to a particular road; it is a challenging task to sort out calls that were carried out while driving on those specific roads. Our study was carried out on a road that has restricted access for movement of pedestrian, animal and bicycle traffic, however, the presence of entry and exit roads still causes unevenness in the traffic flowing along the study segment. Therefore, this will remain as a challenge and a cause for estimation errors.

Despite some of the limitations associated to our study, the analysis in this paper is still interesting by itself and to the best of our knowledge; it is the first time that handovers are used for traffic volume estimation in such complex urban context. We believe the analyses support the claim that handover data can have added-value for traffic volume estimation. Finally, further studies have to be carried out that consider more detailed information on the handovers: lower aggregation than the hour and better information about the orientation of the cell tower sections. One should also bear in mind that the developed model is always inferior to a system that uses a traffic count from loop detector. However, in the absence of traffic data from specific site, engaging a hybrid application that combines data from cellular networks and conventional on-road sensors would provide a better result compared to a solo use of the average city-wide traffic volume as a substitution.

ACKNOWLEDGMENTS

This research was carried out under the framework of the project CityMotion (Data Fusion for Mobility Consumers, Providers, and Planners) within the MIT-Portugal program. We thank

Fundação para a Ciência e a Tecnologia for financing the project. We also thank the project iCIS CENTRO-07-ST24-FEDER-002003 for partially financing the research. We would also like to thank Intergraph Corporation for the use of Geomedia Professional 6.1 (GIS software). Thanks also go to TMN, Portuguese cellphone operator, for letting us use the cellular network data. A final note of appreciation is given to the municipality of Lisbon for providing the traffic counts data that were used in the study.

REFERENCES

Abdulhai, B., Kattan, L., 2004. Traffic engineering analysis in: Kutz, M. (Ed.), Handbook of transportation engineering. McGraw-Hill.

Ahas, R., Aasa, A., Silm, S., Tiru, M., 2010. Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. Transportation Research Part C: Emerging Technologies 18 (1), 45–54.

ANAOCM, 2010. Autoridade nacional de comunicaçoes. Available at: http://www.anacom.pt/render.jsp?contentId=1027764 [Accessed 5 November 2011].

Caceres, N., Wideberg, J.P., Benitez, F.G., 2007. Deriving origin–destination data from a mobile phone network. IET Intelligent Transport Systems 1 (1), 15 - 26.

Caceres, N., Wideberg, J.P., Benitez, F.G., 2008. Review of traffic data estimations extracted from cellular networks. IET Intelligent Transport Systems 2 (3), 179–192.

Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., Ratti, C., 2011. Real-time urban monitoring using cell phones: A case study in Rome. IEEE Transactions on Intelligent Transportation Systems 12 (1), 141-151.

FHWA, 2001. Traffic Monitoring Guide. Office of highway policy information, Washington, D.C.

Herrera, J.C., Work, D.B., Herring, R., Ban, X.J., Jacobson, Q., Bayen, A.M., 2010. Evaluation of traffic data obtained via GPS-enabled mobile phones:The Mobile Century field experiment. Transportation Research Part C: Emerging Technologies 18(4), 568-583.

Hongsakham, W., Pattara-atikom, W., Peachavanish, R., 2008. Estimating road traffic congestion from cellular handoff information using cell-based neural networks and K-means clustering, 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. IEEE, Krabi, Thailand.

Kühne, R.D., 2008. Foundations of Traffic Flow Theory I: Greenshields' Legacy-Highway Traffic, Symposium on the Fundamental Diagram: 75 Years (Greenshields 75 Symposium). Transportation Research Board, Woods Hole Massachusetts, United States.

Leduc, G., 2008. Road traffic data : Collection methods and applications, Working Papers on Energy, Transport and Climate Change N.1. European Comission-Joint Research Centre- Institute for Prospective Technologica, p. 55.

OECD, 2007. Managing urban traffic congestion. European conference of ministers of transport. Available at: http://www.internationaltransportforum.org/Pub/pdf/07Congestion.pdf [Accessed 7 December 2011].

Pan, C., Lu, J., Di, S., Ran, B., 2006. Cellular-Based Data-Extracting Method for Trip Distribution. Transportation Research Record: Journal of the Transportation Research Board(2006-1945), 33-39.

Pulselli, R.M., Romano, P., Ratti, C., Tiezzi, E., 2008. Computing urban mobile land scapes through monitoring population density based on cell-phone chatting. International Journal of design and nature and ecodynamics 3 (2), 121-134.

Ratti, C., Sevtsuk, A., Huang, S., Pailer, R., 2005. Mobile landscapes : Graz in real time, Proceedings of the 3rd Symposium on LBS & TeleCartography, Vienna, Austria

Thajchayapong, S., Pattara-atikom, W., Chadil, N., Mitrpant, C., 2006. Enhanced detection of road traffic congestion areas using cell dwell times, IEEE intelligent transportation systems conference. IEEE, Toronto, Canada.

Thiessenhusen, K.-U., Schäfer, R.-P., Lang, T., 2003. Traffic data from cell phones: a comparison with loops and probe vehicle data.

TRB, 2000. Highway capacity manual. Transportation Research Board, US, Washington, D.C.

Vaccari, A., Liu, L., Biderman, A., Ratti, C., Pereira, F., Oliveirinha, J., Gerber, A., 2009. A holistic framework for the study of urban traces and the profiling of urban processes and dynamics, 12th International IEEE conference on intelligent transportation systems. IEEE, St. Louis, MO.

Valerio, D., Alconzo, A.D., Ricciato, F., Wiedermann, W., 2009. Exploiting cellular networks for road traffic estimation : A survey and a research roadmap, IEEE 69th Vehicular technology conference, Barcelona, Spain.

Varaiya, P., Haoui, A., Kavaler, R., 2008. Wireless magnetic sensors for traffic surveillance. Transportation Research Part C: Emerging Technologies 16(3), 294–306.

White, J., Wells, I., 2002. Extracting origin destination information from mobile phone data, International Conference on Road Transportation and Control, London, pp. 30-34.

Zeng, Q.-A., Agrawal, D.P., 2002. Handoff in Wireless Mobile Networks, in: Stojmenovic, I. (Ed.), Handbook of wireless Networks and Mobile computing. John Wiley & Sons, Inc., pp. 1-26.