Estimating Traffic Speed using Cellular Phone Data

Wang, J.,[1] **Raghavan, V.**[2] **and Song, X. F.**[1]*
[1]University of Chinese Academy of Sciences, Beijing 100049, China
E-mail: wangjing110@mails.ucas.ac.cn, xfsong@ucas.ac.cn
[2]Media Center, Osaka City University, 558－8585 Osaka, Japan, E-mail: raghavan@media.osaka-cu.ac.jp
**Corresponding Author*

Abstract
Traffic speed is an important parameter of traffic flow, particularly during morning or evening rush hours. It could be measured by using inductive loops or cameras, but this is restricted to critical crossroads or streets due to the cost of installations and management. In recent years, using vehicles with cellular phones as probes has become a hot topic to estimate traffic flow on roadways because it takes full advantage of existing cellular network systems to detect the movement of cellular phones. The challenge is to find an effective way to convert the rough positions of cellular phone bearers in cellular systems to exact routes in road network systems. This paper proposes a comprehensive approach that identifies regular commuters by analyzing their historical travel patterns, detects commuting routes using the algorithm of Mixed-Integer Linear Programming (MILP) with constraints, and thereafter estimates current traffic speeds. The approach was implemented using an Open Source GIS solution (Python Plus NetworkX and PostgreSQL) and tested with cellular phone data collected in the city of Haikou, China.

1. Introduction

A cellular (or mobile) network consists of a number of Base Transceiver Stations (BTSs) serving a wide geographic area called a cell. In a city, these stations are often densely distributed, about 300–500 meters away from each other. Cellular (or mobile) phones in the network communicate with other phones, even when their bearers are moving through neighboring cells. Therefore, ubiquitous cellular phones can be treated as low-cost traffic probes or sensors to monitor traffic status without requiring any ancillary devices along roadways (Calabrese et al., 2010 and 2011). In comparison to traditional loop detectors and cameras that are installed on urban freeways, not arterial roads (Rose, 2006 and Bar Gera, 2007), a cellular phone probe is an innovative technique which allows to cover broader areas. A number of references of traffic applications have been published, but some interesting techniques related to the estimation of traffic status are referred to in the following. The early methods tracked a mobile phone by wireless positioning in BTS systems using subsequent signal-strength measurements for different base stations (Hellebrandt et al., 1997 and Hellebrandt and Mathar, 1999). This method greatly increases the load on cellular network systems because it makes big computations and requires additional systems to track all probes. The later methods simply use the default data of cellular phone call detailed records. Astarita and Guido, (2002) proposed a forward propagation mechanism to simulate real-time traffic flow and average speed, in which a road network is discretized into road segments. The concentration of cellular phones at each segment and for each interval of time is calculated to estimate traffic parameters. In addition, the map matching techniques of probabilistic, shortest-distance, or fuzzy-logic based algorithms are also used to match the trajectories of vehicles to their approached roadways for further calculating vehicle speeds (Bar Gera, 2007, Yuan et al., 2010 and Tettamanti et al., 2012). This approach, however, relies heavily on the availability of mobile data in a relatively high spatial resolution. The most recent research studies focused on the statistics of origin-destination (OD) mobility matrices for analyzing the transportation status of road networks. Both Calabrese et al., (2011) and Demissie et al., (2013) inferred traffic counts at given points in the road network by analyzing anonymous phone location data in a mobile phone network. However, the relationship between mobile phone ODs and traffic ODs have not been explored in detail, and Iqbala et al., (2014) solved the problem by an optimization-based approach of weighing and scaling factors. Usually, generating OD patterns for different time periods is based on long-term data collection, supporting traffic planning rather than real-time traffic monitoring. This paper intends to estimate traffic speed using the phone data in cellular network systems. Traffic speed is the average speed of vehicles on a roadway segment over a period of time. We transformed the location of phone bearers in cellular networks into the road networks in urban areas and further estimated their moving speeds on roads. Cellular phone bearers (commuters) who often travel repeatedly between locations were first identified. Most commuters access the routes in the morning and evening every weekday. These routes (consisting of road segments) were detected using a numerical optimization-based route matching algorithm. Finally, the speed of bearers on each road segment during their approaching time were calculated and further weighed to achieve traffic speed on roads on which multiple cellular phone bearers were traveling during the same period.

2. Methods

2.1 Conceptual Model

2.1.1 Characteristics of cellular phone data

Cellular phone tracking data is a time series of data collected in a cellular network system by purposive sampling strategies and consists of system logged data related to mobile phone telecommunication events of phone calls, short messages, Skype or WeChat chatting, and Web browsing. These data include: (1) the Location Area Code (LAC) and Cell Identifier (CID) of base transceiver stations used in an event by a cellular phone at a particular time; (2) the handover time and relevant LACs and CIDs in case the cellular phone approaches the boundary between two LACs; and (3) the power on/off or regular updating information of the cellular phone. Without precise wireless positioning, the positions of cellular phone bearers are approximately presented using the coordinates of base transceiver stations. Therefore, the spatial accuracy of bearer trajectories is of low quality, with an error of 300–500 meters, depending on the density of base stations in urban areas.

2.1.2 Flowcharts of data processing

In this work, the traffic speed along one direction of a road segment of interest is supposed to be the weighed value of the speeds of all vehicles that approach during a given time period. Therefore, it is essential to determine the route or trajectory through which a bearer ever passes and calculates the moving speed of the bearer on each segment. It is not all bearers whose tracking data is related to movement along roadways during a given period. Those bearers with regular commuting routes are preferred because their historical phone data are valuable to increase the confidence in the approach to determining the route. The data processing flowchart is illustrated in Figure 1. In the data process, the space is split into a number of polygons by a Voronoi diagram using the set of points providing locations of the base transceiver stations. Each polygon denotes a cell that is the effective transmission area with respect to a base station.

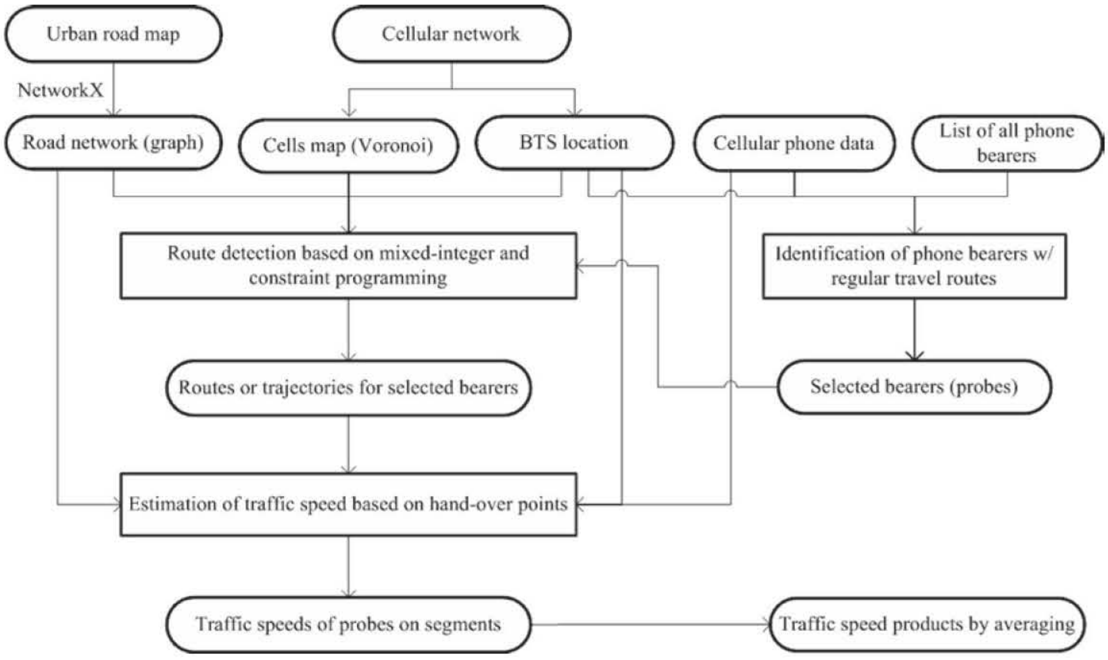


Figure 1: Data processing flowchart

2.2 Identification of Cellular Phone Bearers with Regular Travel Routes

According to statistics, only about 2–5% of cellular phone data are related to the communication events that occur during movement time. Nevertheless, a big number of probes are still moving on roads due to cellular phones having become indispensable personal belongings of humans. In cellular network systems, not all cells (or base stations) that a phone bearer passes through would communicate with the phone. That depends on whether the bearer is communicating using that base station when approaching the relevant cell. That is, the number of active stations (or cells) used in communication is small, although the route of the bearer may overlap a large number of cells. It is very difficult to determine the exact route or trajectory of the cellular phone bearer by the locations of sparse stations. Therefore, historical phone data, i.e., one week's worth, are used here to detect those cellular phone bearers (commuters) who have regular travel routes every weekday and further determine their routes using historical information. To determine whether a phone bearer has a regular travel route, cellular phone data is transformed into a new presentation style: time and distance, as shown in Figure 2. That is, the coordinates of a communication event at a specific time are replaced by the distance of the event location to a reference point. Kernel density estimation (KDE) is adopted to find the places in which a phone bearer dwelled for a long period of time and to further identify the locations as one's place of residence or work by day and night. The kernel function is expressed as a bivariate probability density function, and the kernel density estimator at point *x* represents the sum of the kernel functions as follows (Silverman, 1986):

$$\hat{f}(x)=\frac{3}{nh^2}\sum_{t=1}^{n}K(\frac{1}{h}(x-x_i))$$

Equation 1

$$K(x)=\begin{cases}3\pi^{-1}(1-X^TX)^2, \ X^TX<1\\ 0, \text{ otherwise}\end{cases}$$

Equation 2

Where n is the number of points; h is the bandwidth; x_i is the distance value of the ith point; and $K(x)$ is the kernel function adopted in this work.

2.3 Route Detection by Mixed Integer and Constraint Programming

Once the base transceiver stations used in a communication event by a phone bearer are selected, the exact route can be calculated using constrained Mixed-Integer Linear Programming (MILP) based on a routable road map. In fact, this is a typical elementary shortest path problem with resource constraints. Assume a directed graph $D=(V,A)$, with vertex set V and arc set A, was constructed from the routable road map. The attribute c_{ij} of the arc $(i,j)\in A, i\in V, j\in V$ denotes the length of this arc (road) in this work. $x_{ij}\in\{0,1\}$ indicates whether the arc is selected to compose the shortest path. $V_k, k=1,\ldots,K$ is the set of vertices falling within the kth cell. K is the number of active cells along the route. $\delta^+(i)$ and $\delta^-(i)$ denote the set of forward and backward arcs, respectively, with respect to node i. $A(S)$ is the set of arcs with both ends in $S\subseteq V$. The mathematical model of MILP to determine an elementary shortest path from s to t, meanwhile passing through all the K cells, is defined as follows:

$$\min \sum_{(i,j)\in A} c_{ij}x_{ij}$$

Equation 3

$$\sum_{(i,j)\in\delta^+(i)} x_{ij} - \sum_{(i,j)\in\delta^-(i)} x_{ji} = \begin{cases}1 & if\ i=s\\ -1 & if\ i=t\\ 0 & \text{else}\end{cases}$$

Equation 4

$$\sum_{(i,j)\in A(S)} x_{ij} \leq |S|-1, S\subseteq V_{st}, |S|\geq 2$$

Equation 5

$$\sum_{i\in v_k}\sum_{j\in v-v_k} x_{ij}\geq 1, \ k=1,\ldots,K$$

Equation 6

The objective function (3) is the summary of the costs of the arcs in the shortest path. Constraint (4) ensures a path connecting s and t; constraint (5) is used to eliminate sub-tours or cycles following the Dantzig-Fulkerson-Johnson algorithm; and constraint (6) ensures that the path must overlap each active cell. For constraint (5), a graph with N nodes has 2^N constraints, and this may result in a huge number for a large road network. To reduce the scale of solving problems, the coordinates of base transceiver stations are first chained together, forming a coarse path as shown in Figure 3. All cells that intersect with the coarse path are selected as a buffer zone. The roads that intersect with the buffer zone are the candidate arcs used for MILP analysis.

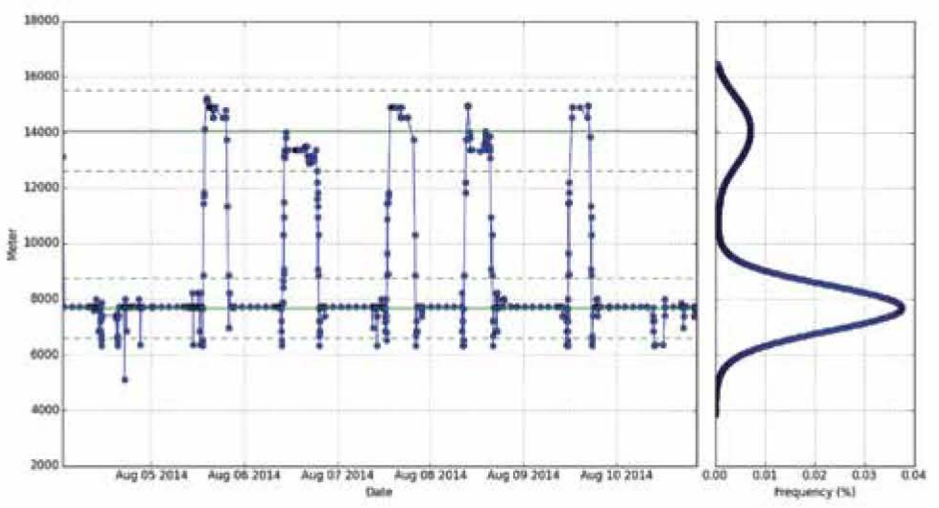


Figure 2: Cellular phone bearer with a regular travel route

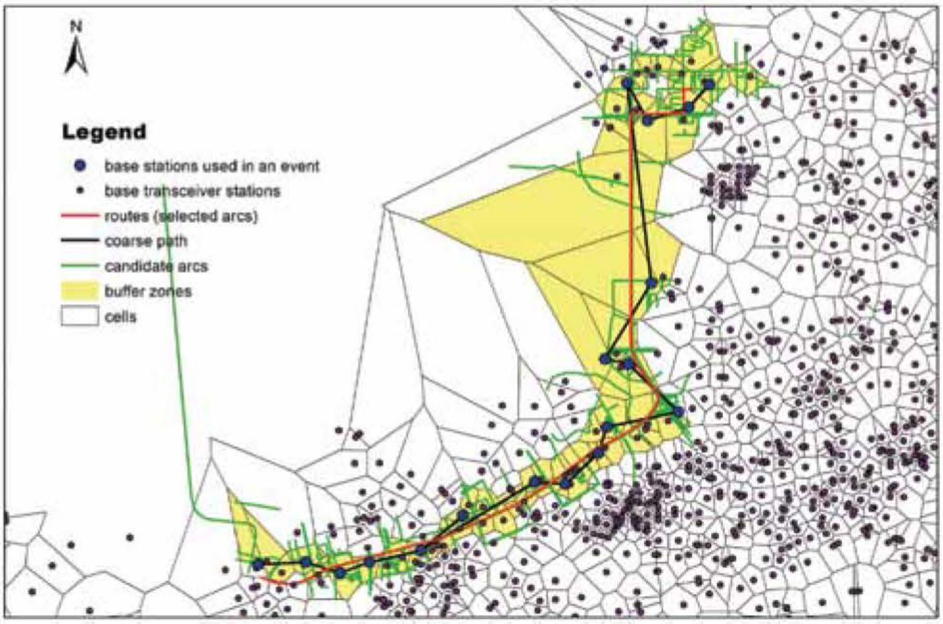
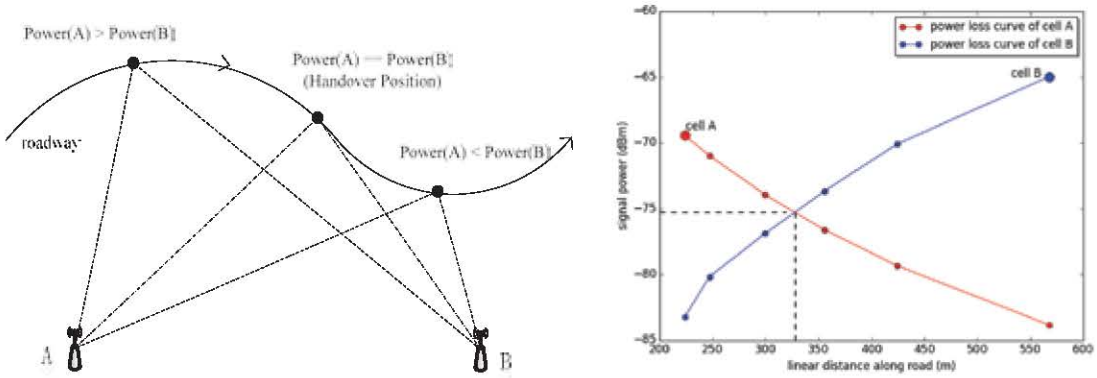


Figure 3: Route detection by MILP



(a) handover position

(b) power loss (attenuation) curves

Figure 4: Determination of handover positions

2.4 Estimation of Traffic Speed Based on Handovers

During communication time, a cellular phone that works with one base station will be switched to another neighbor station when the signal strength received is lower than that of the neighbor or a specific threshold, with respect to the phone bearer's movement along roads. This is called a handover event in cellular network systems. The handover time is recorded in the system, but the position where the handover occurred is not clear. It is often difficult to determine handover points because the transmissions of base stations are affected by complex multiple path reflections in urban areas. A signal power loss (attenuation) method was used to calculate the handover location at which the signal strengths of two neighbor towers are equal on the road. As shown in Figure 4(a), a road passes through the areas dominated by both tower A and tower B, and the handover point, which occurs on the road within their overlapped area, can be determined by finding the point along the road on which the two towers' power is in balance. The field strength of a mobile base station can be calculated using the following equation:

$$P_s = P_t + P_g - L - L_o - P_c$$

Equation 7

where P_s is the signal strength; P_t is the transmit power; P_g is the antenna gain; L is the path loss calculated by the Okumura-Hata model; L_o denotes power losses caused by cable, connector, and instrument, and P_c is the sensitivity of the cellular phone antenna. The Okumura-Hata model is widely used to simulate radio frequency propagation and calculate the power loss of a mobile base station. Transmission between the base station and the cellular phone does have some power loss, known as path loss, which depends particularly on the carrier frequency, antenna height, and distance as shown Figure 4(b). The basic path loss is calculated by the following equations:

$$\begin{aligned} L = {} & 69.55 + 26.16 \lg f - 13.82 \lg h_b - a(h_m) \\ & + (44.9 - 6.55 \lg h_b)(\lg d) \end{aligned}$$

Equation 8

For suburban areas:

$$a(h_m) = (1.1 \lg f - 0.7) h_m - (1.56 \lg f - 0.8)$$

Equation 9

For big urban areas:

$$a(h_m) = \begin{cases} 8.29(\lg 1.54 h_m)^2 - 1.1 \\ (150 < f < 200 MHz) \\ 3.2(\lg 11.75 h_m)^2 - 4.97 \\ (400 < f < 1500 MHz) \end{cases}$$

Equation 10

where L is the path loss (dBm); f is the frequency of transmission (MHz); h_m is the mobile antenna height (meters); h_b is the base antenna height (meters); d is the distance between the base station and phone (kilometers); and $a(h_m)$ is the correction factor of mobile phone antenna height.

3. Results and Discussion

The study area is the city of Haikou, China. There are three telecommunication operators in the city, and we examined cellular phone data (phone calls and short messages) from one operator on 4-10 August 2014 and evaluated the potential of using these data to monitor the traffic speeds of urban road networks. In this period, the users of which the communication event number is more than 150 accounts for about 50%; the event number of 20% users is more than 310, and the maximum value is 1530 times. By statistics, there are about 9.2 million communication events per day, of which about 3% are movement events, and 94% occurred between 8:00 a.m. and midnight, as shown Figure 5. In Haikou, the total length of road networks (including branches) is about 4,716 km, on which 254 movement events occur each minute, for an average of about one probe every 3.6 kilometers within five minutes. Because the average street length is about 600 meters, the number of probes in tested data is not enough to determine traffic speeds within a five-minute interval. Both Web browsing and WeChat chatting data are larger than phone call and message data tested, and therefore, the use of such Internet data or all phone data of the three operators is suggested to significantly increase the number of probes on streets. The data was stored in a database (PostgresSQL), which includes cell information, road network and cellular phone data tables. The algorithms (KDE, Meanshift) and the calculation process (handover, speed) were developed purely in Python language, where two packages were applied. The *NetworkX* was used to generate the graph of road network. The MILP used in this work is a global optimization model that solves the Elementary Shortest Path Problem with Resource Constraints (ESPPRC), which was implemented by ***Openopt***.

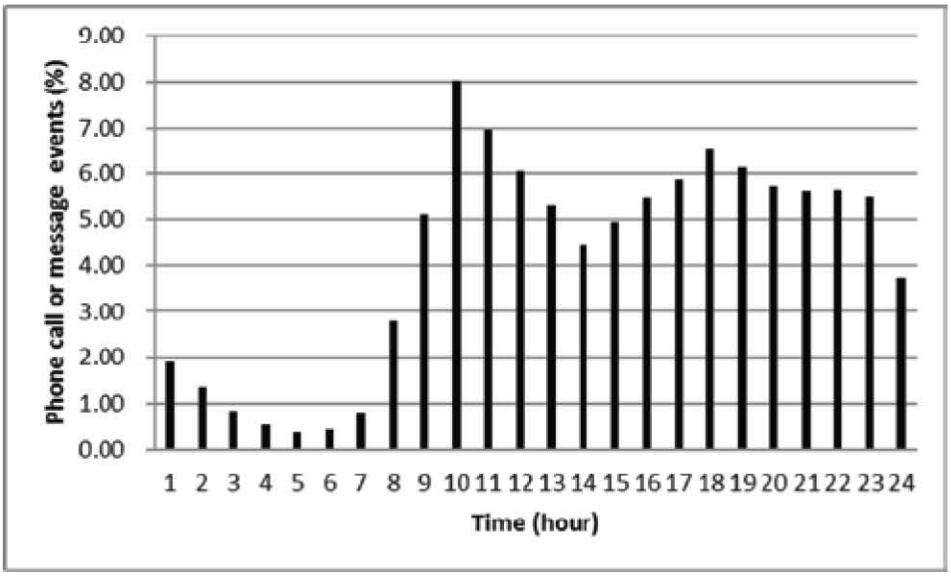


Figure 5: Statistics of cellular phone communication events

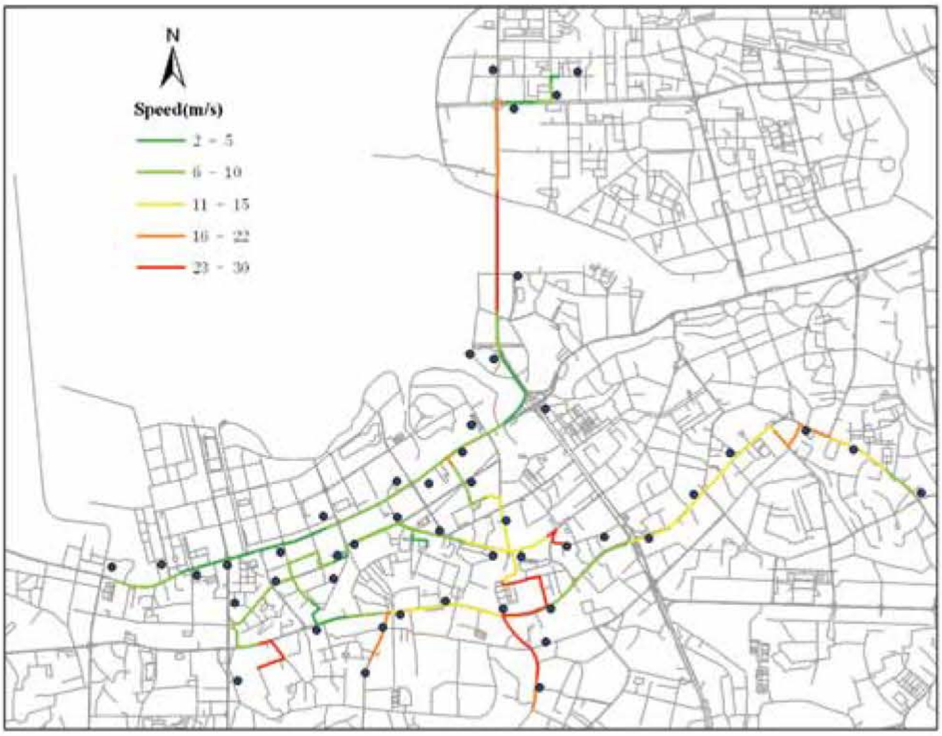


Figure 6: Vehicle traffic speed

It seeks for an elementary shortest path, matching the polygons that a cellular phone bearer approaches. In the resultant route, nodes or arcs cannot be repeatedly visited, which is a limitation or constraint of the ESPP. In this work, the conflicts related to the model limitation were very few because a phone communication event often lasts a short time, i.e., a few minutes, so that the phone bearer could not repeat his/her visit during such a short time even in cases in which he/she was traveling along a loop line. Once there were base stations that were used repeatedly, dynamic programming related to the Djikstra's algorithm was ready to solve the problem. It is time-consuming but is used occasionally. For the computation of vehicle speeds, a road was split into a number of segments by handover points along the road, and the speed at each segment was computed as shown in Figure 6.

Table 1: Differences in handover positions

No.	Estimated (m)		Measured (m)		Difference (m)
	X	Y	X	Y	
1	12942744.39	4852357.63	12942676.71	4852349.75	68.15
2	12942248.44	4852354.50	12942309.98	4852349.81	61.72
3	12941784.89	4852353.44	12941728.37	4852347.95	56.79
4	12941102.29	4852353.88	12941149.95	4852343.94	48.68
5	12940599.81	4852347.14	12940704.26	4852343.84	104.51
6	12939918.53	4852349.18	12939840.39	4852341.18	78.55

Table 2: Differences in speeds

No.	Estimated (m/s)	Measured (m/s)	Difference (m/s)
1	16.71	13.09	3.62
2	15.75	12.31	3.44
3	13.53	16.94	3.41
4	13.22	11.72	1.50
5	7.65	9.70	2.05

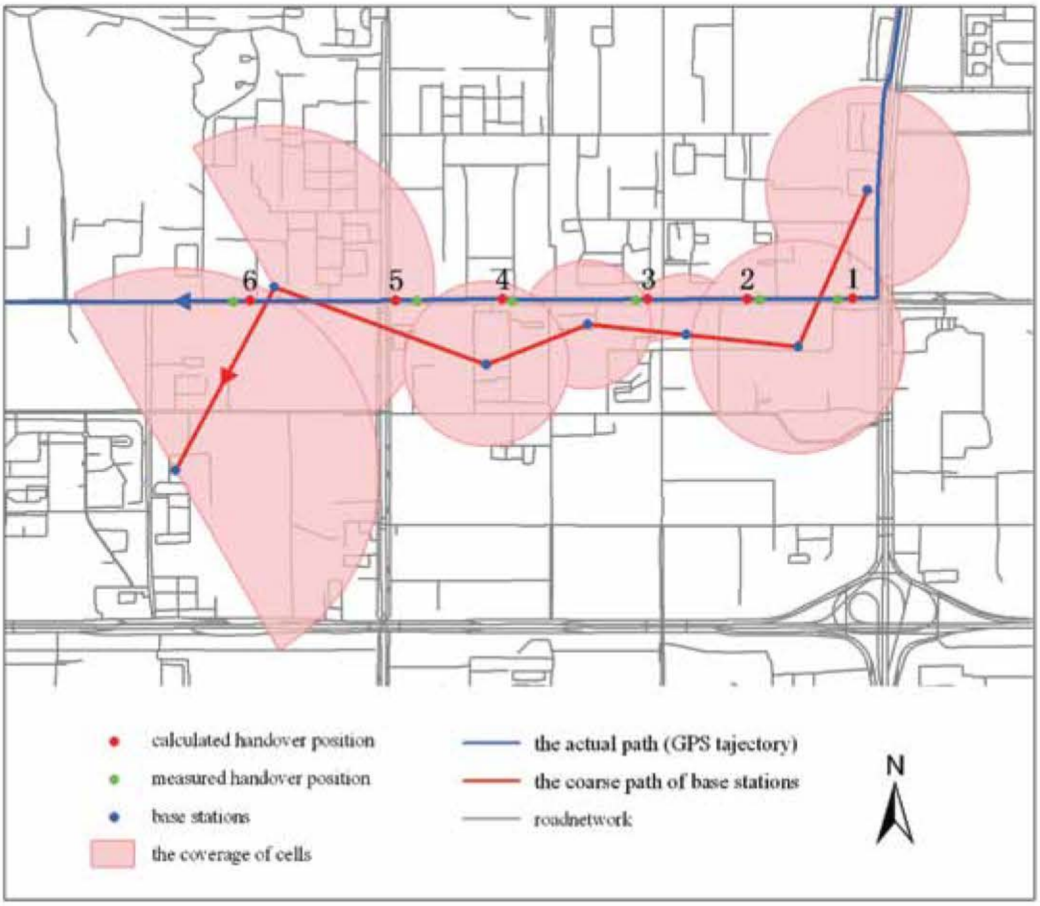


Figure 7: Handover locations along the road (estimated vs. measured)

The traffic speed of roads is the average of the speeds of all segments. The accuracy of the speed is mainly affected by handover points. To evaluate the speed estimation method, we made a ground test along a street by collecting both the power strength data along the road and the location of handover points. As shown in Figure 7, the in-car GPS trajectories were used to compute vehicle speeds and mark the handover locations on the road as ground truth data, and the radio propagation model was adopted to estimate handover positions. We compared estimated handover positions and speeds with those measured. As listed in table 1, the average position difference was 69.73 meters with a standard variance of 18.09. The vehicle speeds among two adjacent handover points are listed in table 2. The mean of the speed difference is 2.80 m/s, and the standard variance is 0.86. The experiment results indicate that using the above method to estimate traffic speeds by cellular phone data is acceptable.

4. Conclusions

This paper proposes a comprehensive approach to estimating the traffic speed of vehicles in urban areas using cellular phone data. Commuters who have regular travel routes between their places of residence and places of work were determined by the kernel density estimation method. The exact travel routes were determined by MILP, in which commuting information increased the confidence in route matching when the commuting route for a cellular phone bearer existed. Vehicle speeds were computed using handover information. The demonstration showed the possibility of using cellular phone data to estimate the speeds of individual vehicles in which passengers communicate during their travel time. To make a high-quality traffic speed product, the multiple resources of cellular phone data are needed, and the estimation of handover positions needs to be modeled using the parameters of the base transceiver stations.

Acknowledgment

This work was supported by the National Science and Technology Major Project of China (2011ZX05039-004) and the Knowledge Innovation Program of the Chinese Academy of Sciences (KZCX2-EW-QN605).

References

Astarita, V. and Guido, G., 2002, Motorway Traffic Management and Traffic Parameters Estimation from Mobile Phone Counts, *The 13th Mini-EURO Conference*, 2002: 534-538.

Bar Gera, H., 2007, Evaluation of a Cellular Phone-based System for Measurements of Traffic Speeds and Travel Times: A Case Study from Israel, Transportation Research Part C: Emerging Technologies, 15(6), 380-391.

Calabrese, F., Colonna, M., Lovisolo, P., Parata D. and Ratti C., 2010, Real-Time Urban Monitoring using Cell Phones: A Case Study in Rome. *IEEE Transactions on Intelligent Transportation Systems*, 12(1): 141-151.

Calabrese, F., Di Lorenzo, G., Liu, L. and Ratti, C., 2011, Estimating Origin-Destination Flows using Mobile Phone Location Data, *IEEE Pervasive Computing*, 10(4): 36-44.

Demissie, M. G., de Almeida Correia, G. H. and Bento, C., 2013, Intelligent Road Traffic Status Detection System through Cellular Networks Handover Information: An Exploratory Study. *Transportation Research Part C: Emerging Technologies*, 32, 76-88.

Hellebrandt, M., Mathar, R. and Scheibenbogen, M., 1997, Estimating Position and Velocity of Mobiles in Cellular Radio Networks. *IEEE Transaction on Vehicular Technology,* 46(1): 65-71.

Hellebrandt, M. and Mathar, R., 1999, Location Tracking of Mobiles in Cellular Radio Networks. *IEEE Transactions on Vehicular Technology*, 48(5): 1558-1562.

Iqbala, M. S., Choudhurya, C. F., Wang, P. and Gonzálezb M. C., 2014, Development of Origin-Destination Matrices using Mobile Phone Call Data, *Transportation Research Part C: Emerging Technologies*, 40:63-74.

Rose, G., 2006, Mobile Phones as Traffic Probes: Practices, Prospects and Issues, Transport Reviews, 26(3): 275-291.

Silverman, B. W., 1986, Density Estimation for Statistics and Data Analysis. London: Chapman & Hall/CRC.

Tettamanti, T., Demeter, H. and Varga, I., 2012, Route Choice Estimation Based on Cellular Signaling Data, *Acta Polytechnica Hungarica*, 9(4): 207-220.

Yuan, Y., Guan, W. and Qiu, W., 2010, Map Matching of Mobile Probes Based on Handover Location Technology, Networking, Sensing and Control (ICNSC), *2010 International Conference on. IEEE*, 2010: 587-592.