Grzegorz Sabak
Faculty of Cybernetics, Military University of Technology, Warsaw

CELL-ID BASED VEHICLE LOCATION ACCURACY – FIELD TESTS REPORT

Abstract: Monitoring of vehicle position and movement via positioning of communication devices inside those vehicles is one of many options to obtain data on stream of vehicles. This paper presents the results of accuracy analysis for the vehicle positioning within GSM network, via method based on cell identifier (Cell-Id location). The analysis takes different types of roads, such as: highways, motorways, trunk roads, voivodeship roads, large urban areas and smaller towns, into consideration. Data for the analysis has been collected during test drives. Cellular network of one of the leading mobile network operators in Poland has been used
Keywords: vehicle telematics, cell-id location, floating car data, floating cellular data, intelligent transportation systems

1. INTRODUCTION

Nowadays, Intelligent Transportation Systems (ITS) are perceived as one of the most promising way of solving problems that the reality puts in front of different institutions and people responsible for fulfillment of society's transportation needs. Similarly to all information systems, the effectiveness of ITS operations depends on the availability and quality of input data, the acquisition of which is a crucial factor in the costs of the ITS system performance.

One of the alternatives (or maybe rather a supplement) for traditional methods of vehicle data collection (such as cameras, induction loops and other) are the methods of data collection via electronic devices used in those vehicles (*Floating Car Data* [1]). At present, however, not all vehicles are equipped with devices that would allow the transmission of data to the central system.

During the last few years, both in Poland and all over the world we have observed a significant increase in cell phone users. According to the Polish Office of Electronic Communications (Urząd Komunikacji Elektronicznej) [2], in 2009, 84.9% of Poles used it, 8% of this group declared ownership of two or more cell phones. At the same time, using wireless communication based on cell network for the machine to machine communication (M2M) becomes more popular.

The above data justifies assumption that, in the majority of the cases, a driver or at least one of the passengers carries a cell phone. A cell phone network is able to monitor the position of its users and can become the source of positioning data for the stream of vehicles (*Floating Cellular Data* [3]). As cell phone networks cover large areas (for instance almost 100% of area of European countries) we can expect that they can provide data even from roads that might not be covered by other types of data-collecting detectors due to costs.

2. LOCATION METHODS IN CELLULAR NETWORKS

In order to properly serve its users, a cellular network has to be able to obtain and process some information about their location. For typical services such as: voice calls, Internet access, short message service (SMS) and some others, it is not required for the network to know geographical coordinates of user's position. It is enough to have information about a cell in a range of which a user was recently. There is a standard procedure (called *paging*) which allows 'finding' a user in a network and establishing a connection when required.

However, there is a set of services (called *Location Based Services* – LBS [4]) which rely on cellular network users' location and require information provided as geographical coordinates. The LBS adjust provided content according to users' location thus making this content much more relevant for the clients. Possible additional revenues from such services have been a key driver for developing different positioning methods. The main objective of this development was to increase accuracy of the location.

The most common method which is available in nearly all cellular networks is a *Cell-Id location* method. In this case, a mapping between a cell identifier and a geographical position has to be created by a network operator. This mapping is then used during processing of a location request. An identifier of a cell which is serving a user is translated to user's position. An accuracy of this method highly depends on sizes of cells in a network and varies from several tens of meters to tens of kilometers.

There are other location methods [5][6] which allow greater accuracy but also require much more capital investment in a network infrastructure. High costs are the reason why network operators rarely decide to implement them in their networks. Those methods relay on some technical parameters related to radio signal propagation time, signal to noise ratio and others. Typically they also require a user to be in coverage of three or more cells to be able to determine his location (which is done via a trilateration method).

In this paper the mathematical model of location data and Cell-Id based positioning in a cellular network is presented. Additionally, the results of accuracy calculations using data obtained during test drives are reported and discussed.

3. DATA MODEL

Let *a mobile object* will be an abstract model of the real-world object that moves on the surface of Earth. An example of such mobile object might be a vehicle, a cell phone, a GSM module or a GPS receiver. *A cell* is an area of radio signal range broadcasted by one of the base transceiver stations (BTS). By *a mobile station* we mean a telecommunication equipment which allows its user to communicate (through voice or data connection) with either other users of the cellular network or external systems (e.g. content servers operating in Internet).

Let us define the following:

- $A \subset \Re^2$ - model of area where the cell phone network 'is active',
- $t \in \Re_{\geq 0}$ - time, timestamp,
- $p, q \in A$ - points in the area (a plane), position,
- $d(p,q)$ - euclidean distance between two points $p = \langle x_p, y_p \rangle$ i $q = \langle x_q, y_q \rangle$,

calculated with formula:

$$d(p,q) = \sqrt{(x_q - x_p)^2 + (y_q - y_p)^2} \tag{1}$$

- $O = \{1,2,...,N_O\}$ - a set of mobile objects identifiers,

Let T be a *timestamp stream* defined as an ordered sequence

$$\mathrm{T} = (t_i)_{i=\overline{1,N_T}} \tag{2}$$

where $t_i \in \Re_{\geq 0}$ and $t_i \leq t_{i+1}$. Let *an event* be a pair $e = \langle t, P \rangle$, where $t \in \mathrm{T}$ is a timestamp defining the moment the event took place and P is a list of parameters that define the type of the event. *An event stream* is defined as a sequence of events:

$$\mathrm{E} = (e_i)_{i=\overline{1,N_E}} \tag{3}$$

We shall denote by $prev(\mathrm{E},t)$ a function, the value of which is an event $e \in \mathrm{E}$, which took place just prior the moment $t \in \mathrm{T}$, and by $next(\mathrm{E},t)$ a function that for the moment $t \in \mathrm{T}$, returns an event that took place just after it. Function $prev(\mathrm{E},t)$ is defined for the time interval $\langle t_{1,} + \infty)$, and function $next(\mathrm{E},t)$ for the interval $(-\infty, t_{N_E}\rangle$.

3.2. Location Data Model

The positioning method Θ is a set of actions the result of which assigns each object $o \in O$ a point $p \in A$ that corresponds to its position in moment $t \in T$.

Let us define the following:

- $\alpha : O \times T \to A$ - *position function* that gives each object $o \in O$ a point that corresponds with its real position in moment $t \in T$. Graph of function α shall be called *object o 's trajectory*.
- $\Gamma^{\Theta} : A \to 2^{A}$ - algorithm of positioning via method Θ - a function that assigns to point $p \in A$ a set of points from the set A which are the possible results of method Θ of positioning for objects located at point $p \in A$.
- $\lambda : O \times T \to A$ shall denote each function that assigns to an object $o \in O$ in moment $t \in T$ a point $p \in A$ that is assumed to be object's position.

If, for the positioning method Θ results of function Γ^{Θ} are single element sets (i.e. points), we will call Θ a *deterministic* method of positioning.

3.3. Location Event Stream

Let *a location event* be defined as a triple $l = \langle t, o, p \rangle$, where $t \in \mathrm{T}$ is a timestamp associated with the event, $o \in O$ is an object and $p \in A$ is its postion. The sequence of location events:

$$\mathrm{L} = (l_i)_{i=\overline{1,N_L}} \tag{4}$$

will be called *a location event stream*.

A polygonal chain built by joining consecutive object positions according to L will be called *an object track*.

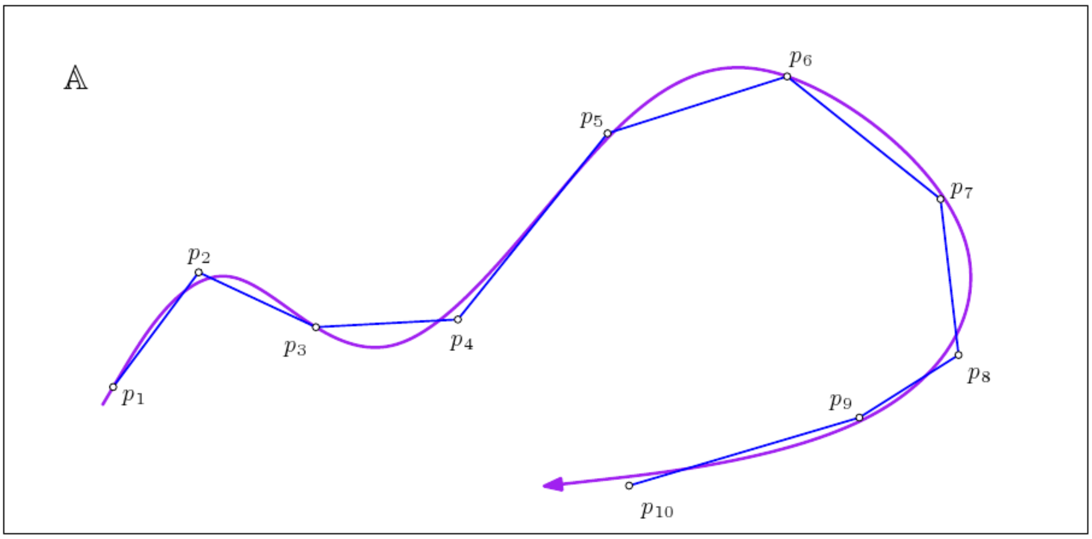


Figure 1. Trajectory and track of an object

Figure 1 shows an example of object trajectory (a smooth curve, arrow indicates the direction of object movement) and track (segmented line) of an object, consisting of ten line segments. In the example above, positioning in some cases does not match the exact position of the object (points: $p_2, p_4, p_5, p_7, p_8, p_9, p_{10}$ do not belong to object's trajectory).

Let $\lambda^{GPS} : O \times T \to A$ denotes localization of the object obtained from the GPS navigation receiver. In the calculations reported in this paper it is assumed that GPS system provides real localization of the object. This assumption can be formulated as:

$$\forall o \in O, \forall t \in T \quad \lambda^{GPS}(o,t) = \alpha(o,t) \tag{5}$$

By $\mathrm{L}^{GPS} = (l_i^{GPS})_{i=\overline{1,N_L}}$ we will denote location event stream obtained with the use of GPS receiver.

3.4. Location Error Stream

Given two location data streams $\mathrm{L}^{\Theta_1}, \mathrm{L}^{\Theta_2}$ obtained via use of location methods Θ_1, Θ_2, *a location error stream* can be defined as a sequence:

$$ERR^{\Theta_1} = (err_i)_{i=\overline{1,N_{ERR}}} \tag{6}$$

where each term is *a postioning error* defined with a formula:

$$err_i = d(\lambda^{\Theta_1}(o_i,t_i), \lambda^{\Theta_2 *}(o_i,t_i)) \tag{7}$$

By $\lambda^{\Theta_2 *} : O \times T \to A$ we denote a function which gives an estimated position of object $o \in O$ in moment $t \in T$ according to data in location stream L^{Θ_2}. An assumption is made that between subsequent location events object was moving with a constant speed and via the shortest way between locations. Performing such estimation is necessary bacuse of fact that typically events in $\mathrm{L}^{\Theta_1}, \mathrm{L}^{\Theta_2}$ are not synchronised.

For each term $l_i^{\Theta_1} = \langle t_i, o_i, p_i \rangle$ of L^{Θ_1}, if both $l_i^< = \langle t_1^<, o_i, p_i^< \rangle = prev(\mathrm{L}^{\Theta_2}, t_i)$ and $l_i^> = \langle t_1^>, o_i, p_i^> \rangle = next(\mathrm{L}^{\Theta_2}, t_i)$ exist, coordinates of point $p_i^* = \langle x_i^*, y_i^* \rangle = \lambda^{\Theta_2 *}(o_i, t_i)$ being an estimation of object's position can be calculated from the formulas:

$$x_i^* = \frac{t_i - t_i^<}{t_i^> - t_i^<}(x_i^> - x_i^<) \tag{8}$$

$$y_i^* = \frac{t_i - t_i^<}{t_i^> - t_i^<}(y_i^> - y_i^<) \tag{9}$$

where $\langle x_i^<, y_i^< \rangle$ are the coordinates of point $p_i^<$ and $\langle x_i^>, y_i^> \rangle$ define point $p_i^>$.

3.5. Cellular Network Model

We shall use the following definitions:

- $B = \{0,1,2,...,N_B\}$ - set of identifiers b of cells that belong to a cellular phone network.

0 means „no cell available".

- $M = \{1,2,...,N_M\}$- a set of mobile stations identifiers, $M \subseteq O$,
- $\theta : M \times T \rightarrow B$ - function which assigns to a mobile station $m \in M$ an identifier of a cell, via which, at the moment $t \in T$, it is being served. If $\theta(m,t) = 0$, it means that in moment t, a mobile station m is not active in the cellular network.

We shall define *a range function* $\gamma : A \times B \rightarrow \{0,1\}$ as a function that assigns value 1 to those points of area A, which are within the cell range, and value 0 to those points that are outside of this range. By the fact of point $p \in A$ being in range of cell b we mean that the mobile station being in point p may be fully served by this cell.

A model of cellular phone network shall be the five:

$$CellNet = \langle A, B, \gamma, M, \theta \rangle \tag{10}$$

Let us define *coverage* of point $p \in A$ as a set of cell identifiers such that:

$$B^p = \{b \in B \ \wedge \gamma(p,b) = 1\} \tag{11}$$

3.6. Location Data in Cellular Network

Let $\beta : B \setminus \{0\} \rightarrow A$ be a function that maps all cell identifiers to points in area A. Point $p^b = \beta(b)$ can be interpreted as a 'center' of the cell b. A positioning algorithm Θ^{CID} based on cell identifier (Cell-Id based location) assigns to a mobile station m a center of the cell b that is serving the mobile station. Cell-Id location function can be defined as following:

$$\lambda^{CID}(m,t) = \beta(\theta(m,t)) \tag{12}$$

The pair $CellNet^{CID} = \langle CellNet, \beta \rangle$ shall be called a model of a cellular network with a Cell-Id based positioning method. A Cell-Id location event stream is a stream of location events in case location was obtained via Θ^{CID} algorithm and will be denoted as $\Lambda^{CID} = (l_i^{CID})_{i=\overline{1,N_L}}$.

4. TESTS RESULTS

In the case of Cell-Id location method, its accuracy depends strongly on a size of a cell within which the mobile station is located. Cell areas vary on different types of terrain (smaller in urban areas, larger in the country and suburbs), which is a direct result of cell network design process in network operating company. It is caused by the fact that cell network operators, when building the infrastructure, try to achieve its adequate quality and capacity, and at the same time they try to optimize investment costs.

In order to assess accuracy of Cell-Id location method a number of tests drives on different types of road were performed between December 2009 and March 2010. The roads where divided into five different categories according to their type. Table 1 lists those categories and provides data which roads where actually analyzed, and how many data samples where collected during test drives (column $\sum N_L$).

Table 1.

Roads analyzed during test drives

Road/area type	Analyzed roads	$\sum N_L$
'highway'	Highways: A1: Gdańsk – Nowe Marzy, A2: Stryków – Poznań, A4: Katowice –Kraków	2 942
'major'	Major roads: DK7: Warszawa – Gdańsk. DK7: Warszawa – Kraków. DK8: Express roads: S6 – Trójmiasto ringroad	19 018
'regular'	Voivodeship roads: 543: Grudziądz – Brodnica, 560: Brodnica – Sierpc, 938: Pawłowice - Cieszyn	5 072
'city'	Streets of Warsaw Centre	1 069
'town'	Streets of smaller towns: Mińsk Mazowiecki, Brodnica	1 977

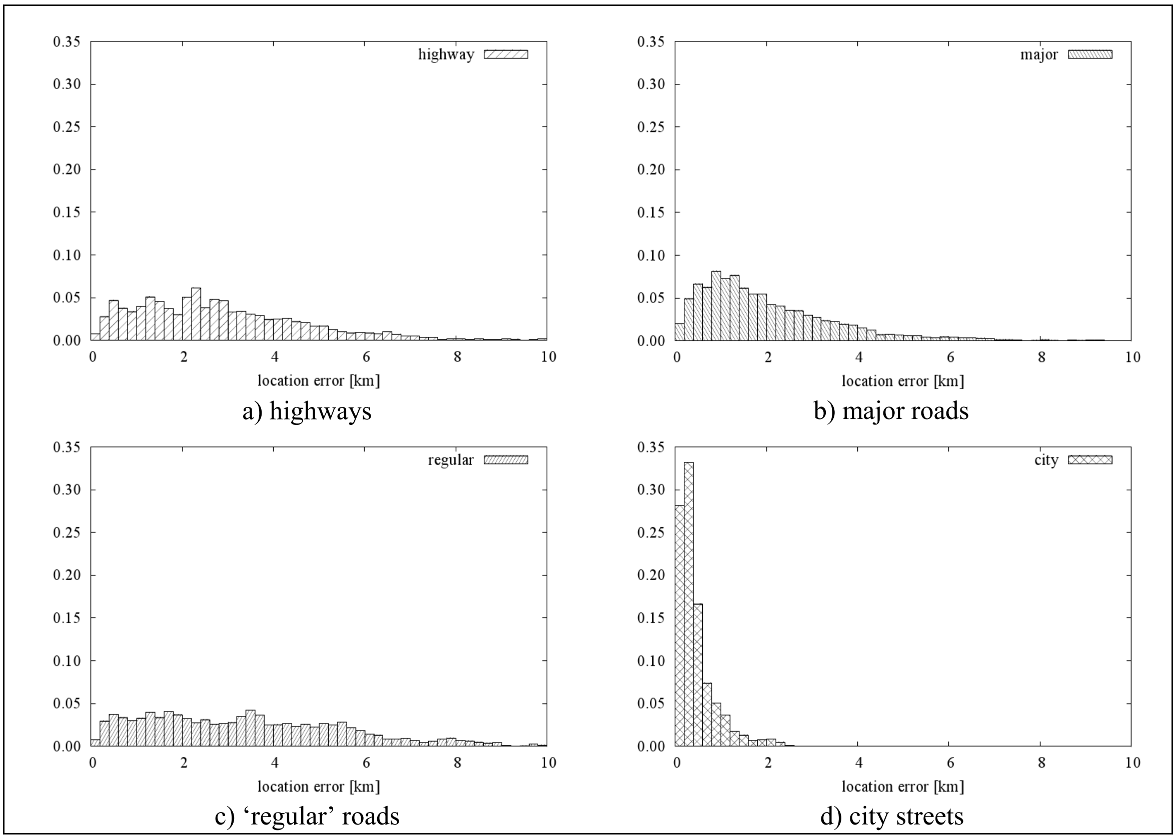
During each test drive, vehicle locations obtained from a GPS receiver and via a Cell-Id method were logged in separate files. Logged data were treated as a pair of location data streams ($\mathrm{L}^{CID}, \mathrm{L}^{GPS}$) based on which an error data stream ERR^{CID} can be calculated. Terms of such error streams were analyzed statistically. The results of this analysis are presented in Table 2. The following statistical data has been calculated: average value, standard deviation, first quartile (Q_{25}), median (Q_{50}), third quartile (Q_{75}) and maximal error (max.).

Table 2.

Statistics of location errors (values in km)

Road/area type	avg.	st. dev.	Q_{25}	Q_{50}	Q_{75}	max.
‘highway’	2,92	2,02	1,42	2,56	3,97	20,37
‘major’	2,66	2,37	1,47	2,09	3,45	27,21
‘regular’	3,47	2,38	1,62	3,23	4,96	28,83
‘city’	0,44	0,41	0,18	0,32	0,53	2,40
‘town’	1,23	1,06	0,51	0,94	1,56	7,77

It can be concluded, that in all cases, the positioning error is characterized by large dispersion. The best positioning accuracy has been observed for road in the centers of large cities. In this case, 25% of measurements had an error smaller than 0.18km and 75% smaller than 0.53km. The worst accuracy has been measured for voivodeship roads (25% below 1.62km, 75% below 4.96km). It is also worth noticing that maximal error values are, for all cases, many times bigger than the median

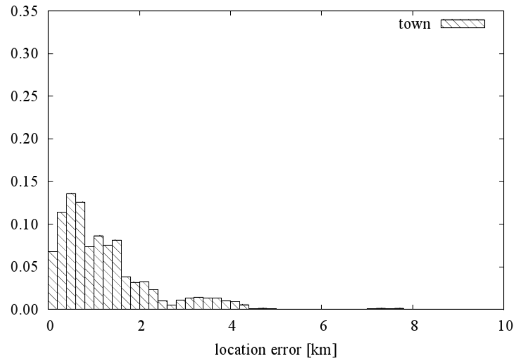


a) highways

b) major roads

c) ‘regular’ roads

d) city streets



e) town streets

Figure 2. Distributions of location errors for analyzed road types

On Figure 2 distributions of location errors for different road types are presented. The charts show similarity between distribution in case of 'highways' and 'regular'. Distributions in case of 'city' and 'town' roads are more concentrated than for roads which are not in urban areas.

5. CONCLUSIONS

Using information on vehicle positioning, based on the cell phone network, is an interesting alternative to all presently used, more expensive sources of data. Location accuracy, however, is one of crucial factors influencing quality of input data and system's operation. Performed tests show that for the Cell-Id location the accuracy is (in case of Poland) on average between 0,44 (city streets) and 3,47 (ordinary roads) kilometers. However, it has to be taken into consideration that in case of non-urban roads location error is characterized by high statistical dispersion. Presented results provide information from the real world cellular network operating in Poland. This data could be effectively used both by scientists focusing on a problems of Floating Cellular Data and vehicle telematics and by decision makers dealing with project aiming to introduce an ITS system on particular areas.

References

1. Leduc G.: Road traffic data: Collection methods and applications. Joint Research Centre, Seville 2008
2. Rynek telekomunikacyjny w Polsce. Klienci indywidualni, Raport PBS DGA na zlecenie UKE, Sopot 2009
3. Fraser S. The use of Floating Cellular Telephone Data for real-time transportation incident management. McMaster University, Hamilton 2007
4. Schiller J.H, Voisard A. Location-based services. Morgan Kaufmann Publisher, San Francisco 2004
5. Michalski W., Przegląd metod określania lokalizacji abonentów w ruchomych publicznych sieciach komórkowych GSM/UMTS z uwzględnieniem dokładności dostarczanej informacji, technicznych możliwości wdrożenia oraz czynników ekonomicznych i prawnych. Etap 1: Charakterystyka metod służących do określania lokalizacji abonentów w sieciach GSM i UMTS", Instytut Łączności, Warszawa 2007

6. CGALIES final report, Report on implementation issues related to access to location information by emergency services (E112) in the European Union", 2002

DOKŁADNOŚĆ LOKALIZACJI POJAZDÓW METODĄ CELL-ID – REZULTATY POMIARÓW TESTOWYCH

.

Streszczenie: Monitorowanie lokalizacji pojazdów poprzez lokalizację zainstalowanych w nich urządzeń do komunikacji w sieciach telefonii komórkowej jest jedną z opcji uzyskiwania danych o strumieniu pojazdów. W artykule przedstawione zostały wyniki analizy dokładności lokalizacji pojazdów w sieci GSM za pomocą metody opartej o identyfikator komórki obsługującej komunikację z urządzeniem (stacją mobilną). Analiza uwzględnia podział dróg i obszarów na typy takie jak: autostrady, drogi krajowe i ekspresowe, drogi wojewódzkie, obszary dużych miast i miast powiatowych. Dane do analizy zostały zebrane podczas przejazdów testowych z wykorzystaniem stacji mobilnych działających w sieci jednego z wiodących operatorów sieci komórkowej w Polsce

Słowa kluczowe: telematyka transportu, dane strumienia pojazdów, dokładność lokalizacji cell-id, inteligentne systemy informacyjne