


RESEARCH

Open Access



# Outdoor location tracking of mobile devices in cellular networks

Jens Trogh<sup>1\*</sup> , David Plets<sup>1</sup>, Erik Surewaard<sup>2</sup>, Mathias Spiessens<sup>2</sup>, Mathias Versichele<sup>3</sup>, Luc Martens<sup>1</sup> and Wout Joseph<sup>1</sup>

## Abstract

This paper presents a technique and experimental validation for anonymous outdoor location tracking of all users residing on a mobile cellular network. The proposed technique does not require any intervention or cooperation on the mobile side but runs completely on the network side, which is useful to automatically monitor traffic, estimate population movements, or detect criminal activity. The proposed technique exploits the topology of a mobile cellular network, enriched open map data, mode of transportation, and advanced route filtering. Current tracking algorithms for cellular networks are validated in optimal or controlled environments on a small dataset or are merely validated by simulations. In this work, validation data consisting of millions of parallel location estimations from over a million users are collected and processed in real time, in cooperation with a major network operator in Belgium. Experiments are conducted in urban and rural environments near Ghent and Antwerp, with trajectories on foot, by bike, and by car, in the months May and September 2017. It is shown that the mode of transportation, smartphone usage, and environment impact the accuracy and that the proposed AMT location tracking algorithm is more robust and outperforms existing techniques with relative improvements up to 88%. Best performances were obtained in urban environments with median accuracies up to 112 m.

**Keywords:** Outdoor, Positioning, Localization, Location-based services, Mobile cellular network, Real-time, Filtering, Big data, Spark

## 1 Introduction

Network-based positioning algorithms locate a mobile user based on measured radio signals from base stations in its vicinity. The growing amount of available positioning data has led to many location-based services (LBS). These are a collection of applications that use geographical location data of mobile devices provided by Wi-Fi, Bluetooth Low Energy (BLE), Global Positioning System (GPS), or cellular networks [1]. They provide services for end users, e.g., wayfinding in large shopping centers or hospitals, personal navigation, and location-based gaming. This is also important for businesses and government, e.g., asset tracking, fleet management, optimizing productivity in manufacturing or distribution, analyzing traffic patterns, transportation planning, security, and surveillance [2, 3]. A more specific example is the estimation of population

movements during disasters or outbreaks. These require timely and accurate location data which large-scale surveys cannot provide, whereas network operators manage data which can potentially be used to calculate location data in real time [4].

The main contribution of this paper is the novel positioning algorithm: AMT (antenna, map, and timing information-based tracking) to accurately locate all mobile users in a cellular network without any required modifications at the mobile side (client) or network side (server). The latter is useful for applications where there is typically no cooperation at the mobile side, e.g., traffic monitoring, population movement estimation, or criminal activity detection. The proposed location tracking algorithm exploits enriched open map data [5], a mode of transportation estimator, and advanced route filtering on top of the mobile cellular topology and measurements to track the movement and locations of mobile devices. Furthermore, it does not depend on additional or custom software, forced messages, dedicated infrastructures,

\*Correspondence: [jens.trogh@ugent.be](mailto:jens.trogh@ugent.be)

<sup>1</sup>Department of Information Technology, IMEC - Ghent University, Ghent Belgium

Full list of author information is available at the end of the article

direct communication between mobile users, or prior training data.

An extensive experimental validation was conducted that included trajectories on foot, by bike, and by car, in urban and rural environments while a person was actively using his or her smartphone, but also in standby mode. In this mode, all applications that use the mobile network are blocked (e.g., email and messaging services) and as such, standby mode represents a worst case scenario in terms of the number of location updates. The latter shows measurement gaps of up to 6 min while a user was on the move, i.e., time periods where no measurement data is available, which mainly occur in rural areas. Current existing location tracking algorithms for mobile cellular networks are not able to cope with large measurement gaps but instead are deployed in optimal or controlled environments with a high base station density, regularly available measurement updates, large training sets, or are merely validated by simulations with a fixed location update rate. The novel contributions of this paper are:

- An immediately applicable location tracking algorithm that does not require any modifications to the client or network side
- The algorithm does not depend upon any prior training via, e.g., offline fingerprinting, drive-testing, or crowd-sourced measurement campaigns
- Confirmed to work for a large set of users, nationwide, and in real time based on an experimental validation instead of merely relying on simulations

The paper is structured as follows. Section 2 describes the related work. Section 3 outlines the mobile network and grid configuration, type of measurements, and trajectories for the experimental validation. Section 4 discusses the proposed location tracking algorithm in detail, and Section 5 presents the results. Finally, in Section 6, conclusions are provided.

## 2 Related work

### 2.1 GPS enabled

The Global Positioning System (GPS) is a satellite-based navigation technique that is ubiquitous due to its widespread use and worldwide coverage. It can be used to track mobile devices but only if the GPS receiver is enabled, the location data is transmitted to a central server, and there are no GPS outages. The latter refers to the unavailability of GPS signals from sufficient satellites due to, e.g., mountains, tall buildings, or multi-level overpasses. Possible solutions are geometry-based location techniques [6]. A system that utilizes mathematical geometry to estimate vehicle location focusing on road trajectory and vehicle dynamics is presented in [7].

### 2.2 Infrastructure enabled

The most widely used approach to locate a mobile device with telecommunication data from a network infrastructure is cell-ID based [1]. The mobile user is mapped to the location of its serving base station, i.e., the cell to which a mobile device is currently connected. It has a low cost and a short response time and is easy to implement and applicable in all places with cellular coverage but has a low accuracy for high cell ranges.

The most common signal parameters used for network-based location tracking are angle of arrival (AoA) [8], time of arrival (ToA) [9], time difference of arrival (TDoA) [10], and amplitude (signal strength) [11]. AoA techniques determine the direction of propagation of a radio frequency wave and require an antenna array at the side of the incoming wave (network side). This technique performs especially well in line-of-sight (LoS) conditions. ToA techniques measure the time the radio signals travel between a single transmitter (mobile user) and multiple receivers (base stations), and requires two-way ranging or synchronization between transmitter and receiver [12]. In TDoA techniques, time differences between the time of flight of multiple radio signals are measured at the receiving base stations; this is used in, e.g., LoRa [13]. Amplitude-based techniques convert the received signal strength to a distance based upon a path loss (PL) model for distance conversion; however, it is required that an accurate PL model is known for the considered environment. Knowledge of the network topology to estimate the distances between a mobile user and a set of base stations reduces the positioning for all signal parameters to a triangulation or multilateration problem [14]. Sensor fusion techniques combine two or more of these signal parameters to estimate the location [15].

Alternatively for the amplitude-based technique, the location can be estimated by searching for the closest match in a fingerprint database or coverage map. This look-up table maps possible positions with a vector of associated signal strength values or cell-IDs from a set of base stations [16]. The signal strength values are collected in an offline phase and can be measurement-based by test-driving the area of interest [17–19], simulation-based by using a propagation model [20–22], ray tracing [23], or a hybrid approach [24]. Drive-testing is labor intensive and needs to be redone each time the mobile network or even the environment undergoes changes. Also, possible locations for the mobile user are limited to places where a car can pass, meaning no indoor, pedestrian, or off-road locations will be estimated. The simulation-based approach is much faster but will generally lead to less accurate location estimations. Alternatively, a crowd-sourced measurement campaign can be used instead of drive-testing.

Network-based location tracking poses several problems due to multipath and non-line-of-sight (NLoS) conditions, small-scale and large-scale fading, low signal-to-noise ratios, and interference by other mobile users. These affect the radio signal parameters used as input data to location tracking algorithms. To process the noisy signal parameters and improve the accuracy, location tracking algorithms use additional intelligence and information. NLoS mitigation techniques use more robust estimators or simply discard the NLoS component [9]. Map-based algorithms use information about the environment to limit possible locations and transitions between two location updates; this can be done in combination with Kalman filters [25], particle filters [18], hidden Markov models (HMM) [26], data fusion [27], or least squares estimator [28]. A database correlation technique over Received Signal Strength Indication (RSSI) data that is based on advanced map- and mobility-based filtering is presented in [29]. The algorithm is validated in a field environment with trips by car, a location update rate forced to 2 Hz, and an electromagnetic field simulator. A cooperative positioning technique for cellular systems using RF pattern matching is presented in [30]. It is shown in simulations that leveraging the device-to-device (D2D) communication protocol can improve positioning performance if insufficient base stations are visible to a user entity. A crowd-sourced measurement campaign to develop radio frequency (RF) coverage maps and a similarity-based location algorithm is presented in [31]. A proprietary application, installed on the smartphone of a sample set of users in the network, periodically reports the RF channel measurement along with the GPS tag to a central server, which are then processed into the RF coverage map. This resulted in accuracies up to 50 m and 300 m, depending on the cell's coverage range. A semi- and unsupervised learning technique that minimizes the effort to label signal strength measurements for the network-side cellular localization problem is presented in [32]. This technique uses Gaussian mixture models to model the signal strength vectors and an expectation maximization approach to learn the distributions. Accuracies up to 30 m are reported as long as enough training data is available and the base station density is high. A machine learning technique for indoor-outdoor classification and particle filter with HMM for cellular localization is presented in [18]. The trajectory of a moving user was synthesized and reconstructed based on a data training set of around 129000 drive test data points and a fixed location update interval of 10 s, which led to accuracies up to 20 m in urban environments. Note that the latter accuracies are only achieved with large (crowd-sourced) training sets, synthesized data, and high location update rates, which our approach does not require. Furthermore, the proposed location tracking algorithm is

confirmed to execute in real time for more than a million users in parallel and outperforms state-of-the-art particle filters [18].

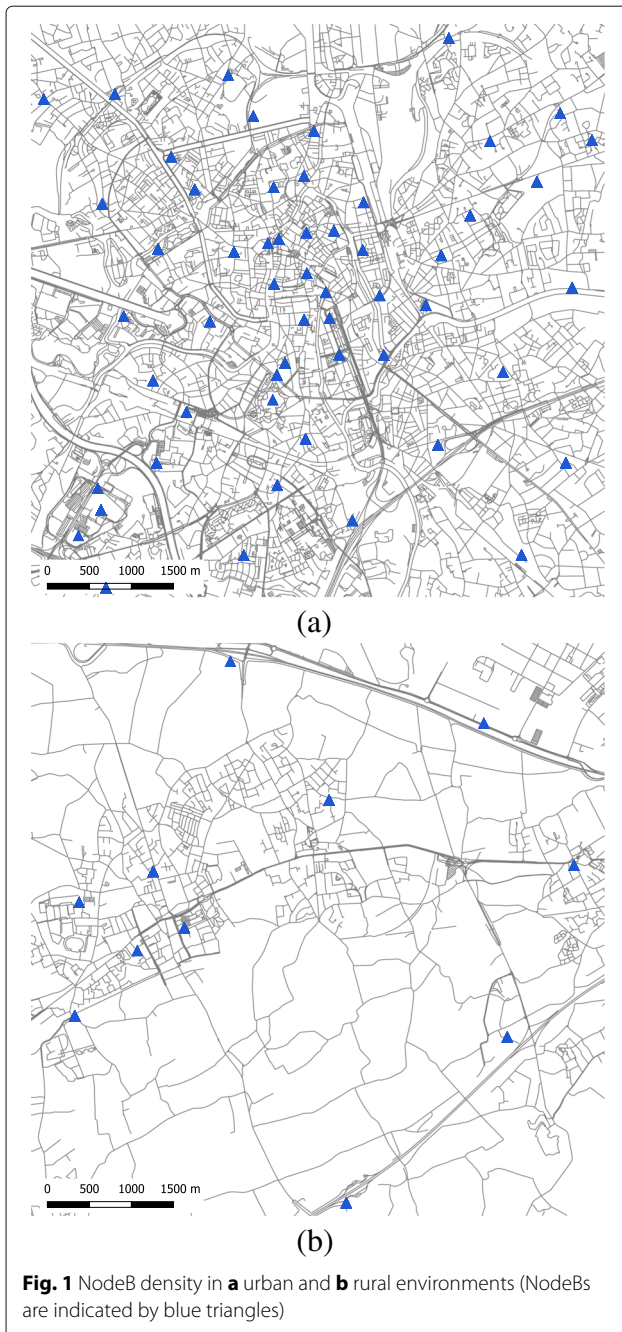
A topic which currently attracts a lot of attention is user anonymity. Mobile network operators ensure anonymity between their mobile users by providing a temporary identifier (TMSI) instead of constantly using the long-term unique identifiers (IMSI). Lately, also anonymized location data has become a subject of concern [33, 34]. Countermeasures to tackle these exposed vulnerabilities are proposed in [35, 36]. In [37], simulations are used to calculate the number of devices necessary to locate non-participant individuals in urban environments. They prove that it is possible to track the movement of a significant portion of the population with a high granularity over long periods of time when a small part of the population is part of a (malicious) sensor network.

### 3 Methods

#### 3.1 Network

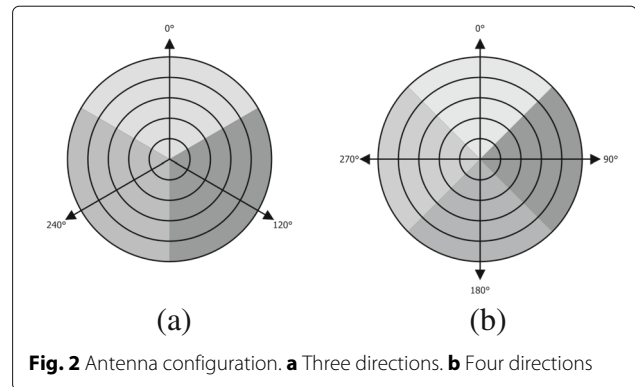
The mobile network, which is used in the experimental validation, consists of more than 2500 NodeBs (September 2017), distributed over Belgium's territory (30528 km<sup>2</sup>). In a 3G network, the base stations are referred to as NodeBs. Figure 1 shows the NodeB locations in a representative urban and rural environment on the same scale (i.e., Ghent and Melsele respectively). It is clear that the environment will have a major influence on the positioning accuracy, because of the difference in NodeB density on the one hand and in urban planning on the other: a sparser road network can limit plausible locations, and the type and height of buildings can affect the signal parameters used as input to location tracking algorithms (e.g., apartments vs. stand-alone houses vs. office buildings). There are more than 50 NodeBs in an area of approximately 45 km<sup>2</sup> for the experiments in an urban environment, whereas for the rural environment there are roughly 10 NodeBs in an area of the same size. The comparison and influence on the performance are discussed in Section 5.

A NodeB has multiple antennas with unique cell-IDs, oriented towards different directions (Fig. 2). Antenna configurations with one up to six distinct orientations occur in the mobile network, which is used in the experimental validation, the most common ones are with three (92%), one (4%), and two (3%) different antenna directions. Usually, a mobile user will connect to the NodeB antenna that is directed towards him. Likewise, a user for which measurements are available from antennas with different orientations but from the same NodeB has a large chance to be located between both zones. The aforementioned observations provide information that is exploited in the proposed location tracking algorithm (Section 4).



### 3.2 Grid

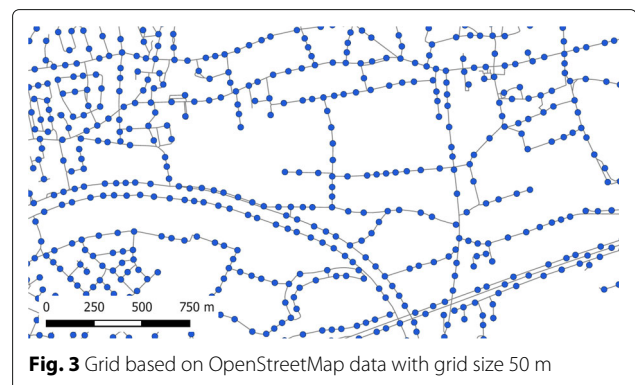
The grid represents a collection of points in the area of interest where a mobile user can be located. In a regular Cartesian grid, all elements are unit squares. It is a simplistic approach where all areas are equally important and take the same resources in both database size and processing time. Alternatively, a map-based grid can be used to limit possible points along the major (motorway, freeway, primary, secondary, and tertiary) and minor (local and residential) roads from the area of interest. The grid



size determines the number and density of these points. Our grid is based on OpenStreetMap data, which consists of straight line segments enriched with metadata about the type of road, information about one-way traffic, relative layering, street name, and maximum allowed speed. Every start point and endpoint of a straight road segment is automatically included in the grid, and road segments are further divided into pieces equal to the grid size. The dots in Fig. 3 represent such a grid with grid size 50 m. For Belgium, this results in 3.2 million grid points for the map-based technique instead of 12.2 million for a Cartesian grid.

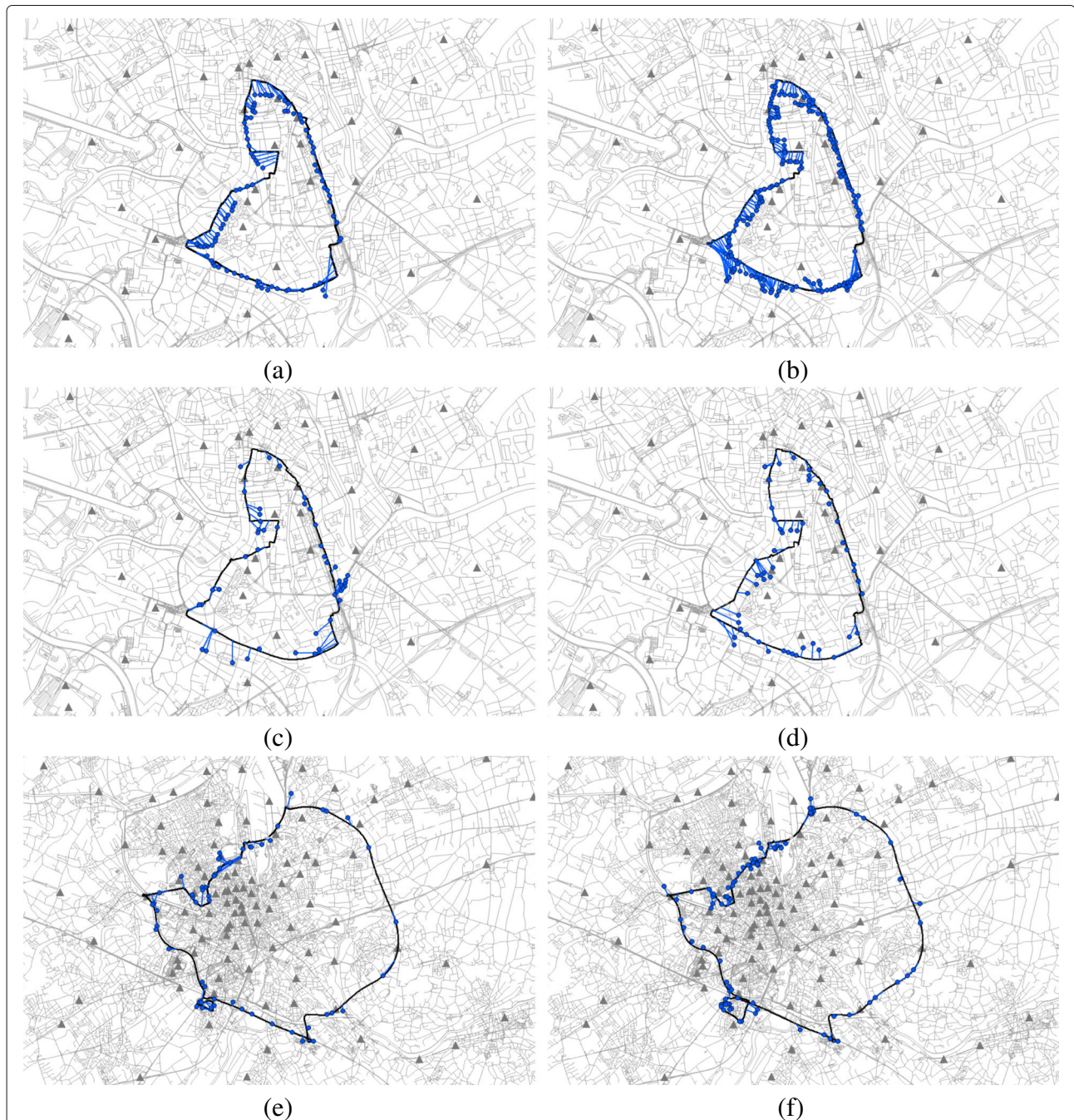
### 3.3 Experiments

Experimental data is collected in cooperation with a major network operator in Belgium. The experiments are conducted in and around the city center of Ghent and in a smaller town near Antwerp (Melsele), to represent urban and rural environments, and on the highway between both cities. The mobile network collects 3G data for more than a million mobile subscribers but to quantify the location errors (accuracy), the real position or ground truth needs to be known, for which permission and cooperation of a mobile user are needed. The experimental validation encompasses trajectories on foot, by bike, and by car. A

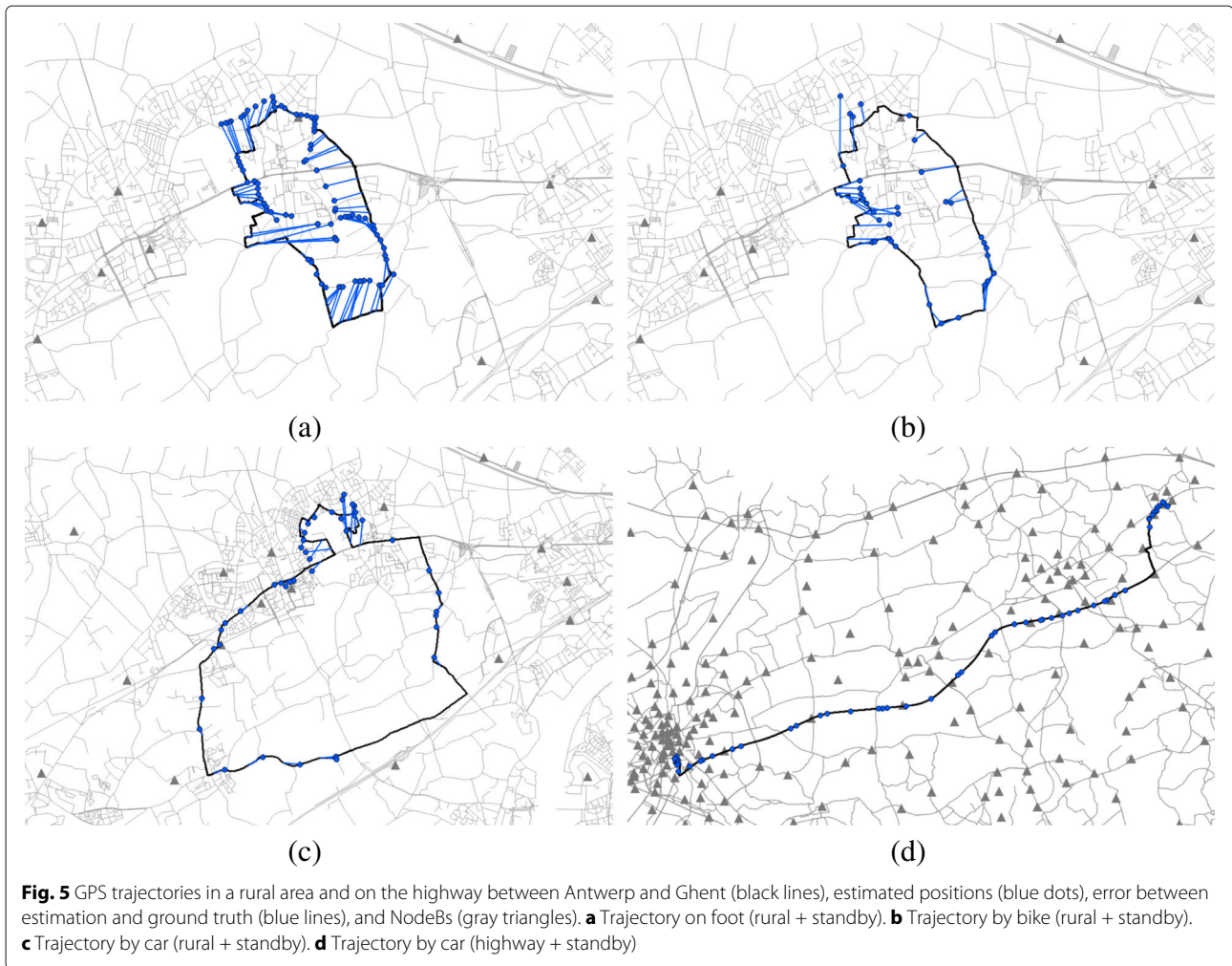


smartphone with a GPS logging application is carried in all scenarios by a mobile user. It was put in the dashboard holder for the car rides and carried in the pocket for the trajectories on foot and by bike. The smartphone was forced on 3G to make the experiments independent of having 4G coverage and to ensure a fair comparison between urban and rural environments. Figures 4 and 5

show the GPS trajectories as black lines, and the sample rate of the GPS logging application was set to 1 location per second. The NodeB locations are indicated with gray triangles. The GPS trajectories are post-processed with a map matching algorithm [38] to increase the accuracy; this is especially useful in urban areas near tall buildings (urban canyoning). Section 4 describes the location



**Fig. 4** GPS trajectories in and around the city center of Ghent (black lines), estimated positions (blue dots), error between estimation and ground truth (blue lines), and NodeBs (gray triangles). **a** Trajectory on foot (urban + standby). **b** Trajectory on foot (urban + streaming). **c** Trajectory by bike (urban + standby). **d** Trajectory by bike (urban + streaming). **e** Trajectory by car (urban + standby). **f** Trajectory by car (urban + streaming)



tracking algorithm, and Section 5 discusses the performance and accuracy for all trajectories in detail. The total distance, duration, and average speed for all trajectories are summarized in Table 1.

### 3.4 Measurement data format

3G measurements are made by the mobile network, i.e., by the radio network controller that controls the NodeBs.

**Table 1** Trajectory details

Scenario	Distance [km]	Duration [min]	Average speed [km/h]
Walk (urban)	8	84	6
Bicycle tour (urban)	8	25	19
Car ride (urban)	39	25	47
Walk (rural)	8	101	5
Bicycle tour (rural)	8	22	22
Car ride (rural)	19	28	41
Car ride (highway)	48	36	80

The input data for our location tracking algorithm are timing information and received signal strength values from a set of NodeBs. Both are reported on regular time periods but independently from each other. The timing information comes in the form of a *propagation delay* and is reported only by the serving NodeB. The signal strength values originate from the *measurement reports* and are reported for all NodeBs that a mobile device currently sees (i.e., from which it receives a broadcast message). Timing information to these other NodeBs would require network changes and increases the load in the mobile network and, hence, is not used in our approach.

#### 3.4.1 Propagation delay

The *propagation delay* parameter can be used to estimate the distance between a mobile device and its serving cell. This delay is used by the radio network controller to make communication possible. It checks and adjusts this delay to allow transmission and reception synchronization. The propagation delay has a time granularity of 780 ns, which corresponds to 234 m [39]. A value of 1 means the mobile

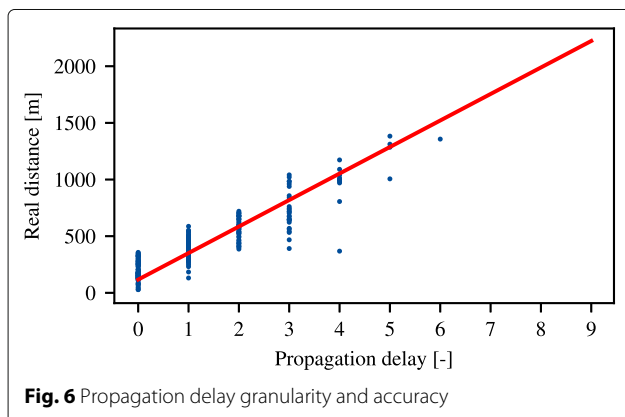
user is located in the interval between 234 and 468 m from the NodeB, from which we derive the following formula to convert propagation delays to distances:

$$\text{distance} = 234 \cdot (\text{propagation\_delay} + 0.5) \quad (1)$$

Figure 6 shows a plot of the real distance (between mobile user and NodeBs) as a function of the observed propagation delay parameter, during a walk of 8 km in the city center of Ghent, Belgium (Fig. 3b). For this test, a radio application was installed on the mobile device and was permanently streaming audio to ensure regular network updates and measurement data. The walk took 84 min during which 234 propagation delay measurements with 49 different cell-IDs from 15 NodeBs were recorded (one physical NodeB can have multiple cell-IDs depending on the number of supported frequencies and different orientations of its antennas). The maximum measured propagation delay during this walk in the city center of Ghent was 6, which corresponds to 1521 m. In rural areas, propagation delays up to 22 ( $\approx 5$  km) were recorded with the same mobile device, which is to be expected due to the sparser base station density. The measured propagation delays fall in the correct interval in 69% of the observations. They are one, two, and three units apart in 27%, 3%, and 0.4% of the cases, respectively. The mean and standard deviation of the absolute differences between the real and calculated distance are 94 m and 82 m. These values are to be expected with a distance granularity of 234 m (i.e., the calculated distances, based on the 3G propagation delays, are in steps of 234 m). Note that the proposed technique can also be applied on 4G and 5G measurements, which have a higher base station density and more accurate timing information, and therefore, will yield a higher location precision (e.g., 4G has a time granularity of 260 ns, corresponding to 78 m).

### 3.4.2 Measurement report

Measurement reports contain information about channel quality and are reported by a user entity (mobile device)



to a NodeB. They assist the network in making handover and power control decisions. The received signal code power (RSCP) denotes the power measured by a mobile user on a particular physical communication channel, also known as common pilot channel. It continuously broadcasts the scrambling code from the NodeBs and carries no other information. These broadcast messages are transmitted with a constant transmit power and gain but can differ per NodeB (information available in network topology). The measurement reports contain measured signal strength values from all NodeBs the mobile user currently sees. As such, the RSCP values can be converted to a path loss value:

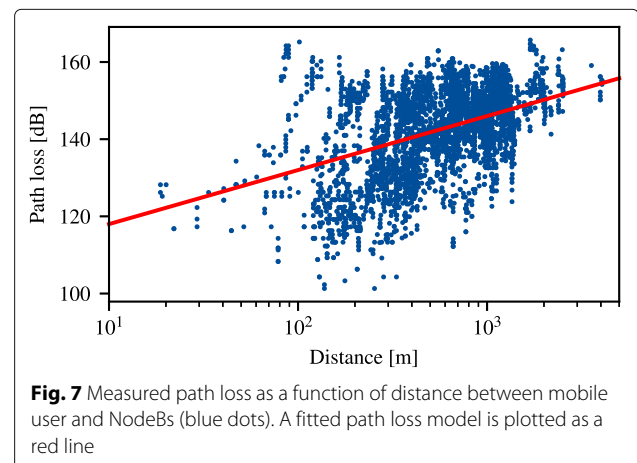
$$PL = P_{TX} + G_{TX} - RSCP \quad (2)$$

where  $PL$  [dB] denotes the total path loss,  $P_{TX}$  [dB] and  $G_{TX}$  [dB] are the transmit power and gain of a NodeB, respectively, and  $RSCP$  is the received signal strength code power measured by a mobile device.

Figure 7 shows these path loss values on the  $y$ -axis and associated distances between mobile user and NodeBs on the  $x$ -axis (the measurement reports are collected in the same experiment as the propagation delays from Section 3.4.1). During the experiment, 578 measurement reports were collected with 4106 RSCP values to 136 different cell-IDs from 32 NodeBs. The fitted one-slope path loss model (red line) has the following form:

$$PL = PL_0 + 10\gamma \log_{10} \left( \frac{d}{d_0} \right) + X_\sigma \quad (3)$$

where  $PL$  [dB] denotes the total path loss,  $PL_0$  [dB] is the path loss at a reference distance  $d_0$  [m],  $\gamma$  [-] is the path loss exponent,  $d$  [m] is the distance along the path between transmitter and receiver, and  $X_\sigma$  [dB] is a log-normally distributed variable with zero mean and standard deviation  $\sigma$ , corresponding to the large-scale shadow fading. The measurement data from this experiment yields a  $PL_0$  of 118 dB at a reference distance of

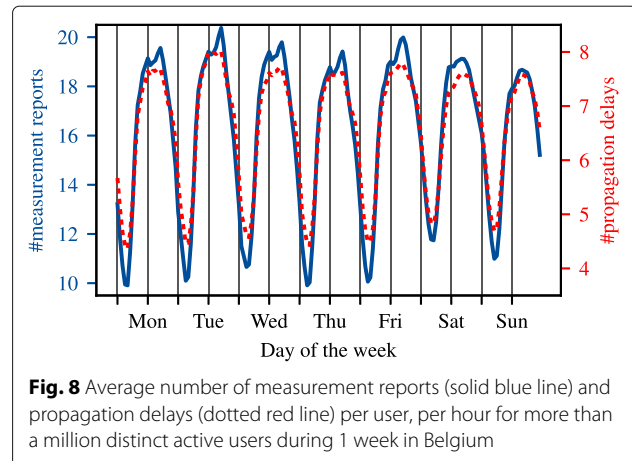


10 m with a  $\gamma$  of 1.40, resulting in an  $R$ -squared of 23% and a standard deviation of 9.8 dB. Low  $R$ -squared values indicate that the data is not close to the fitted line, which results in bad estimations. Also, deviations in measured path loss will result in larger errors at greater distances to the NodeBs, e.g., for a deviation of 5 dB: a value of 135 dB (164 m) instead of 140 dB (372 m) results in a location error of 209 m and a value of 145 dB (848 m) instead of 150 dB (1931 m) results in an error of 1082 m. These larger errors occur rather often, and 26% of the measurements have a user to NodeB distance that is greater than 1 km. As such, the mean and median absolute errors for all 4106 measured values are 1143 m and 473 m, respectively. These location errors are much higher compared to those derived from the propagation delay, suggesting that many received path loss measurements contain no additional information and can worsen the accuracy when used together with the timing information as input to a location tracking algorithm. Note that these path loss measurements can be useful in combination with fingerprint maps based on test-driving or crowd-sourced measurement campaigns but these are labor intensive or require modifications on the client side [16, 18, 19]. Also, these crowd-sourced measurement campaigns will be heavily influenced by, e.g., passing cars, new buildings, or other infrastructure changes.

### 3.5 Cellular network data

The problem with cellular network data is the limited amount of available data, which determines the number of possible updates. Mobile devices can support a range of different wireless technologies, e.g., infrared, Bluetooth, Wi-Fi, GPS, Universal Mobile Telecommunications System (UMTS) in 3G networks, and Long-Term Evolution (LTE) in 4G systems, but not all data are available to the network operator and this also depends on the usage of a mobile user.

Figure 8 shows the average number of measurement reports and propagation delays, per user, per hour, during one week, measured on a 3G mobile network in Belgium for more than a million distinct active users. It is immediately clear that every day exhibits a similar pattern for both the measurement reports and propagation delays with the difference that there are about twice as many measurement reports. The least and most active hours are 3 a.m. and 6 p.m. respectively ( $x$ -axis ticks are set every 12 h and the labels are set at 12 p.m.). Saturdays and Sundays show a flatter and lower curve than weekdays because more people are staying at home, which translates into fewer measurements per user on the mobile network during the day. On Friday and Saturday between 11 p.m. and 5 a.m., there is an average increase of 20% in number of measurements for a similar amount of people compared to weekdays, indicating that there is more movement or



**Fig. 8** Average number of measurement reports (solid blue line) and propagation delays (dotted red line) per user, per hour for more than a million distinct active users during 1 week in Belgium

usage of mobile devices (whether or not outdoors). There are more than a million distinct active users during the whole week, but the maximum number of active users in 1 h is only 700k. This is because not all users send updates to their mobile network when he or she is not moving, has WiFi coverage, or is on a different mobile network (2G or 4G). Current time-series or map-based tracking algorithms assume regular measurement updates to filter outliers and improve the accuracy [25, 29]. This assumption does not hold for many mobile users, making the aforementioned algorithms not generally applicable. The proposed location tracking algorithm can cope with this and consists of multiple phases, depending on the amount of available measurements. Also, it is successfully validated, in cooperation with a major network operator in Belgium, to work in real time on more than a million subscribers with an Apache Spark implementation to support fast cluster computing. The used cluster consists of nine nodes with a total memory of 1.58 TB and 408 physical cores.

## 4 Location tracking algorithms

The performance of the proposed location tracking algorithm will be compared with two reference algorithms: cell-ID (Section 4.1) and centroid based (Section 4.2). The new tracking algorithm is presented in Section 4.3.

### 4.1 Cell-ID

The first reference algorithm is the most simplistic, where a mobile user is mapped to the NodeB to which it is currently connected (also known as serving NodeB or serving cell-ID). This approach is easy to implement and has a low cost and short response time but usually has the lowest accuracy [40].



## 4.2 Centroid

The second reference algorithm takes all different NodeBs from the measurement reports into account and calculates the centroid. In case there is only one NodeB with measurements, this approach results in the same location as the cell-ID technique. Alternatively, a weighted centroid algorithm can be used, where NodeBs get a weight assigned based on their measurement frequency or received signal strength information [41].

## 4.3 AMT: antenna, map, and timing information-based tracking

Figure 9 shows a flow graph of our proposed location tracking algorithm, which uses the orientation of NodeB antennas, map, and timing information as input (AMT). Phase I processes the data measured by the radio network controllers and calculates the temporary estimations (TEs). Phase II further refines these estimated locations with a route mapping filter that uses OpenStreetMap (meta)data, measurements from a recent past (user history), and an estimated mode of transportation as input.

### 4.3.1 Phase I: temporary estimation

The pseudo-code to calculate the temporary estimation of a user, residing on the mobile cellular network, is shown in Algorithm 1 and the variables and steps are discussed with an example in the text below.

Consider the example in Fig. 10: a mobile user is located in the center (yellow square), its serving NodeB ( $cell_{sc}$ ) is indicated with a green star ( $loc_{sc}$ ), and there are three

### Algorithm 1: Calculating the temporary estimation.

---

**Data:** measurements reported by radio network controller

**Result:** temporary estimation (TE)

$cell_{sc} \leftarrow$  cell-id of the serving cell

$loc_{sc} \leftarrow$  location of  $cell_{sc}$

$\alpha_{sc} \leftarrow$  antenna orientation of  $cell_{sc}$

$\beta_{sc} \leftarrow$  opening angle of  $cell_{sc}$

$pd \leftarrow$  reported propagation delay

$CA_{sc} \leftarrow$  area bounded by circular arcs based on  $loc_{sc}$ ,  $\alpha_{sc}$ ,  $\beta_{sc}$ , and  $pd$

$circleSectors \leftarrow$  empty list

$JoinedMeasReport \leftarrow$  measurement reports where  $\alpha$  en  $\beta$  are joined if same cell location // explained in text

**for**  $MR \in JoinedMeasReport$  **do**

$cell_{nb} \leftarrow$  cell-id of NodeB that reported  $MR$

$loc_{nb} \leftarrow$  location of  $cell_{nb}$

$\alpha_{nb} \leftarrow$  antenna orientation of  $cell_{nb}$

$\beta_{nb} \leftarrow$  opening angle of  $cell_{nb}$

$CS_{nb} \leftarrow$  circle sector based on  $loc_{nb}$ ,  $\alpha_{nb}$ , and  $\beta_{nb}$

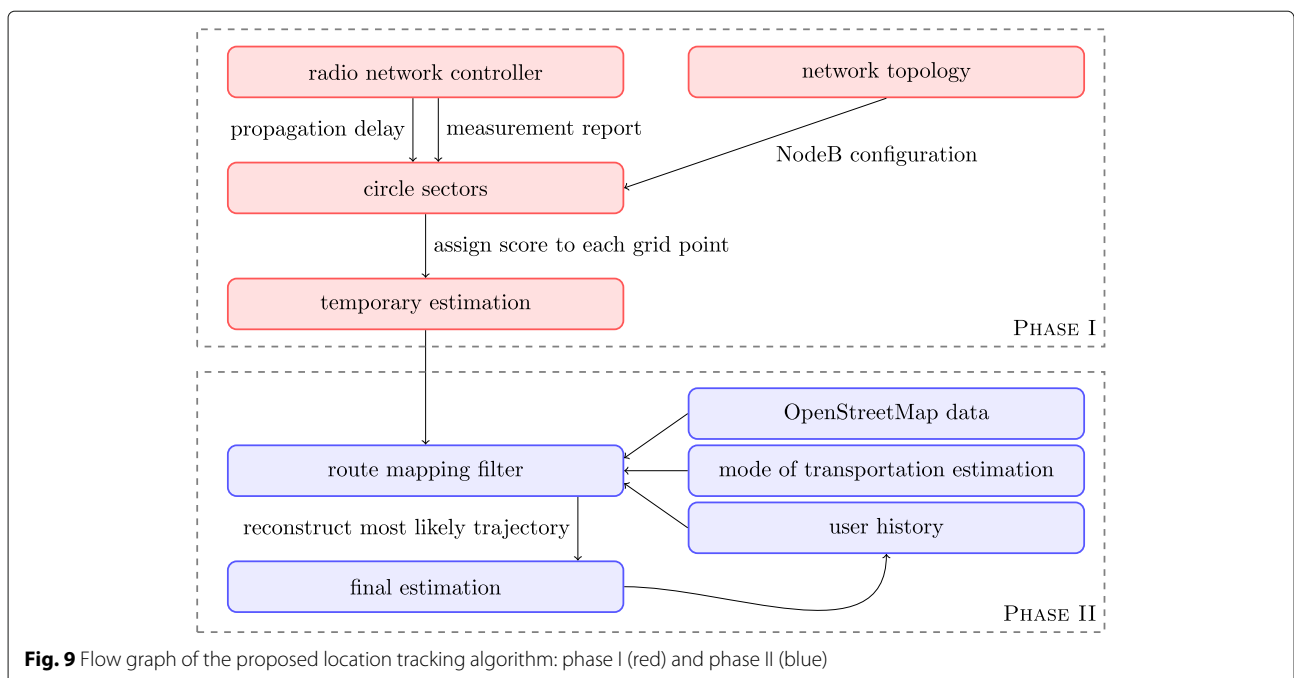
add grid points within  $CS_{nb}$  to  $circleSectors$

**end**

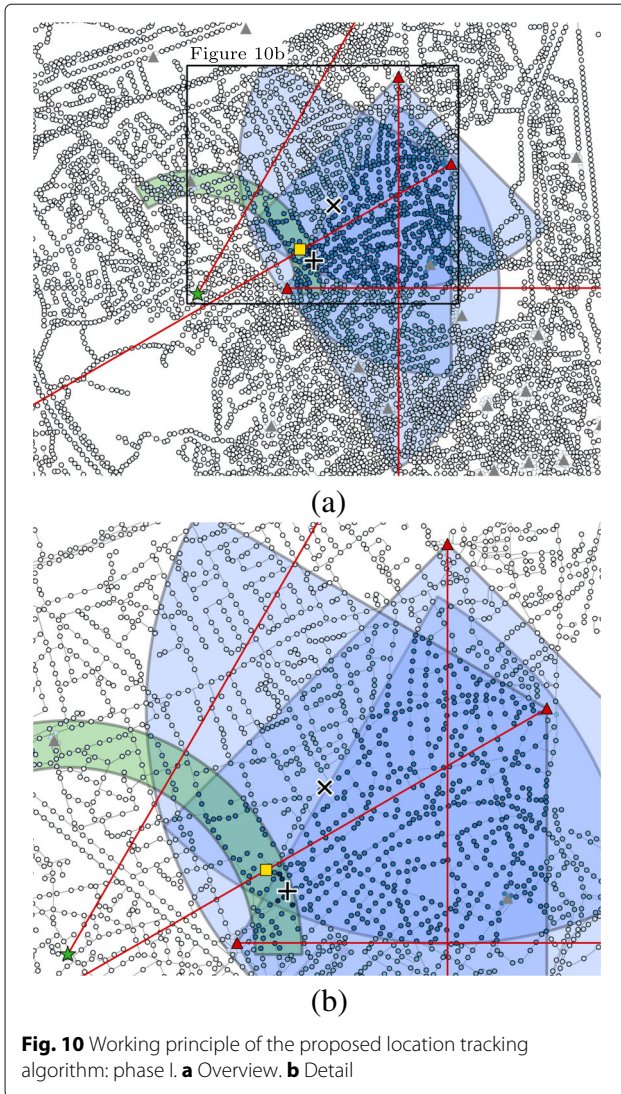
$GPMO \leftarrow$  grid points within  $CA_{sc}$  with maximum occurrences in  $circleSectors$

$TE \leftarrow$  median location of  $GPMO$

---



**Fig. 9** Flow graph of the proposed location tracking algorithm: phase I (red) and phase II (blue)



other NodeBs ( $cell_{nb}$ ) for which there are signal strength measurements ( $loc_{nb}$  indicated with red triangles). The antenna orientations ( $\alpha_{sc}$  and  $\alpha_{nb}$ ) of cell-IDs with measurements are indicated with a red line. The other NodeBs in this area (without measurements at this time instance) are shown as gray triangles, and the grid points are shown as regular dots on top of the road network. The radio network controller reports a propagation delay of 4 from the serving NodeB, which triggers a new location update. This propagation delay corresponds to 1053 m, which limits the possible locations to an area ( $CA_{sc}$ ) bounded by two circular arcs with an opening angle ( $\beta_{sc}$ ) of  $120^\circ$  and radii of 936 m and 1170 m (indicated in transparent green on the left). The distance between both arcs is based on the time granularity of 3G (780 ns corresponds to 234 m). A window of 5 s is used to link measurement reports with propagation delays since they are not reported at

the exact same time instances. Because the calculated distances based on the reported signal strengths from the measurement reports are not reliable (Section 3.4.2), only the orientation ( $\alpha_{nb}$ ) and opening angle ( $\beta_{nb}$ ) of the antennas corresponding to these measurements are used. These are retrieved by looking up the reported cell-ID in the network topology, resulting in three additional circle sectors ( $CS_{nb}$ ), indicated in transparent blue.

The opening angle of the sectors depends on the number of antennas and different orientations the NodeBs have and is equally divided between all orientations. The most common case of three distinct and equally spread antenna orientations corresponds to an opening angle of  $120^\circ$  (similar to the different gray zones in Fig. 2a). If there are multiple measurements to one NodeB and the reported cell-IDs correspond to antennas with different orientation, then both measurements are merged (*Joined-MeasReport*) and a new circle sector is used instead, i.e., the smallest area between both orientations. For example, if there are measurements received on the antennas with directions  $0^\circ$  and  $90^\circ$ , then the new circle sector would be the first quadrant ( $0$  to  $90^\circ$ ) instead of the area from  $-45$  to  $135^\circ$  (see Fig. 2)b. Because users that are located just outside a circle sector could be picked up by the antenna, as is visible in Fig. 10 for the antenna on the bottom center, a margin of  $10^\circ$  is added to the left and right side of a sector.

The coloring of the grid points corresponds to the number of NodeBs (cell-IDs) that are visible from this grid point (it is visible if a grid point falls within the sector areas defined above). In this case, there are only 6 locations that satisfy all measurements, i.e., inside the propagation delay area and in all three circle sectors (green and blue areas). The median location of this set (*GPMO*) is the temporary estimation, indicated with a black plus sign (+) in Fig. 10.

If there is no overlap between the propagation delay area and the circle sectors (green and blue areas respectively), then the median location of the propagation delay area is used as temporary estimation. The latter happens in only 2% of all location updates in our experimental validation (Section 5). Using this approach results, for the depicted example, in an error of 132 m, whereas the cell-ID approach would map the mobile user to the serving NodeB (indicated with a green star), resulting in an error of 1103 m, and the centroid approach results in an error of 490 m (indicated with a black cross  $\times$ ).

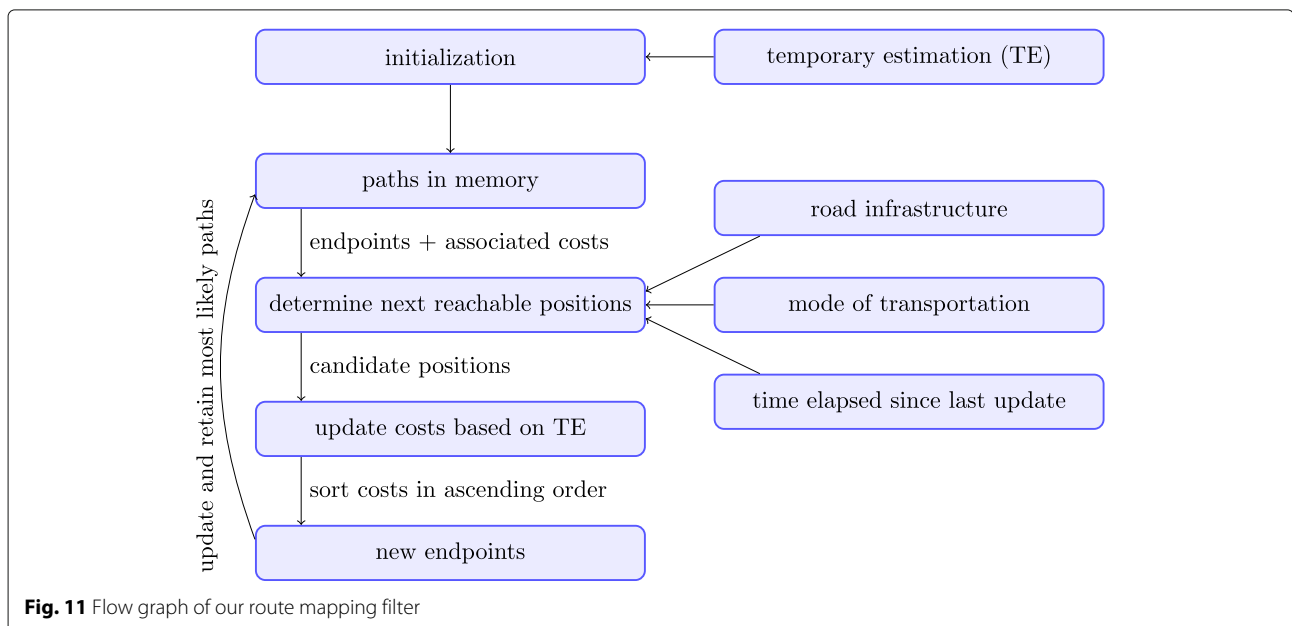
#### 4.3.2 Phase II: route mapping filter

These temporary estimations can be improved with a route mapping filter if there are location updates available from a recent past (user history). For example, a user on foot will travel far less than a user by bike or by car, given a certain time interval. Furthermore, the most likely trajectory over a certain time period can be reconstructed

by making use of OpenStreetMap (meta)data: road infrastructure (ways); maximum speed limits; one-way street information; type of road, e.g., sidewalk, bike path, or highway; and the user's measurement history. To take into account cars that are speeding and to avoid that location estimations are lagging behind, the allowed speed limit (for the reconstructed trajectory) can be increased by, e.g., 10% for each road segment. The proposed route mapping filter is based on the Viterbi path, a technique related to hidden Markov models [21, 22]. By processing all available data at once, previous estimated locations can be corrected by future measurements (similar to backward belief propagation). Naturally, this is only possible if the intended application tolerates a certain delay. A differentiation between real time and non-time critical will be made in the route mapping filter's output. Figure 11 shows a flow graph of our proposed route mapping filter which ensures realistic and physically possible paths.

The pseudo-code of the route mapping filter is shown in Algorithm 2, and the variables and steps are discussed in the text below. For the first positioning update or if there is no location history available from a recent past, the temporary estimation is taken as the current position ( $TE_0$ ). Then, a predefined number of other locations are selected around this position and their cost is initialized to 0, e.g., the 1000 closest grid points to the current position ( $MP$ ). This ensures that the route mapping filter can recover from faulty first estimations, i.e., 1000 grid points and a grid size of 50 m result in a covered surface of roughly 2.5 km<sup>2</sup> (the exact area depends on the road density). The initialization forms the starting point of all

possible paths that are kept in the memory of the location tracking algorithm (*pathsInMem*). Next, when the mobile network reports *new measurements*, a new  $TE$  is calculated as described in Section 4.3.1. After that, for all *paths* in memory, all reachable positions ( $RGP$ : parent grid point) starting from the path's current last grid point ( $PGP$ : parent grid point) are determined by making use of the surrounding road network, the time elapsed since last location update ( $\Delta t$ ), estimated mode of transportation ( $MoT$ ), and OpenStreetMap metadata (maximum speed, type of road, and one-way information). These reachable positions, which are also grid points, are the candidate positions for the next location update. Each candidate position ( $CP$ ) retains a link to the parent grid point ( $PGP$ ) and a cost that represents this new branch along the road network (*pathsTemp*). For time-critical applications, the path which currently has the lowest cost is used as real-time location estimation ( $AMT-RT$ ). In this case, previous estimated locations will not be corrected by future measurements, only the user's current history is taken into account. Lastly, the  $MP$  paths with lowest cost are retained to serve as input for the next iteration when the mobile network reports *new measurements*. At the end of an experiment or measurement interval, all parent grid points from the path with lowest cost are visited in backwards order; this results in the final estimated trajectory:  $AMT-NTC$  (non-time critical). Figure 12 shows a detail of the locations before and after the route mapping filter for the trajectory on foot in Ghent (Fig. 3b). The temporary estimations are indicated with green crosses, and the final estimated trajectory with blue dots.



**Algorithm 2:** Route mapping filter.

---

**Data:** temporary estimations (TE)  
**Result:** most likely trajectory mapped on roads:  
 AMT-RT (real-time) and AMT-NTC  
 (non-time-critical)

$TE_0 \leftarrow$  first temporary estimation  
 $t_{prev} \leftarrow$  first timestamp with measurements  
 $MP \leftarrow 1000$  // maximum paths in memory  
 $pathsInMem \leftarrow$  list with  $MP$  grid points closest to  
 $TE_0$  initialized with cost 0

**while** new measurements **do**

$t \leftarrow$  current timestamp  
 $\Delta t \leftarrow t - t_{prev}$   
 $TE \leftarrow$  current temporary estimation  
 $MoT \leftarrow$  estimated mode of transportation  
 $pathsTemp \leftarrow$  empty list

**for**  $path \in pathsInMem$  **do**

$cost \leftarrow$  cost of  $path$   
 $PGP \leftarrow$  endpoint of  $path$  (parent grid point)  
 $RGP \leftarrow$  reachable grid points (along roads with  
 $MoT$  within time span  $\Delta t$  starting from  $PGP$ )  
 // calculate new path cost for  
 each candidate position (CP)

**for**  $CP \in RGP$  **do**

$dist \leftarrow \sqrt{(CP_x - TE_x)^2 + (CP_y - TE_y)^2}$   
 $path_{new} \leftarrow path + CP$   
 $cost_{new} \leftarrow cost + dist$   
 add  $(path_{new}, cost_{new})$  to  $pathsTemp$

**end**

**end**

$pathsInMem \leftarrow$  retain  $MP$  paths from  $pathsTemp$   
 based on lowest cost  
 // current most likely position at  
 time  $t$   
 $AMT-RT_t \leftarrow$  endpoint of  $path$  with lowest cost  
 $t_{prev} \leftarrow t$

**end**

// most likely trajectory given all  
 measurements  
 $AMT-NTC \leftarrow path$  with lowest cost in  $pathsInMem$

---

The maximum allowed speed used by the route mapping filter can be refined if the mode of transportation is correctly estimated, e.g., pedestrians or cyclists will usually not move faster than 6 km/h or 30 km/h respectively. In our approach, the mode of transportation is estimated based on the rate and distance between serving cell handover zones, i.e., when a new NodeB becomes the serving cell. When a handover takes place, the middle between both NodeBs (estimated handover location) is saved together with the timestamp the handover took

place. The average speed between all estimated handover locations that took place during a certain moving window is used to label the mode of transportation. A moving window of 10 min (5 min before and after the location update) could be used for the non-time-critical route mapping filter, but this is not possible for real-time applications (as no future measurements are available). For this reason, only the last 5 min (counting backwards from the location update that is being calculated) is considered to estimate the average speed. It is labeled as walking if it is below 10 km/h, as cycling if it is between 10 and 25 km/h, and otherwise as driving a motorized vehicle. In the latter case, the route mapping filter will continue to use the maximum allowed road speed for each segment. Although, the location updates (TEs) are more frequent and accurate than the estimated handover locations, they show more fluctuations which results in an overestimation of the average speed (see Fig. 12). For example, during the walk in the city center of Ghent (Fig. 3b), there are 232 location updates whereas there are only 48 handovers, which result in an average estimated speed of 25 km/h based on the location updates and 7 km/h based on the estimated handover locations with a moving window of 5 min.

## 5 Results and discussion

### 5.1 General

Figures 4 and 5 show the estimated positions with the proposed location tracking algorithm as blue dots. The errors between the GPS ground truth and estimated positions are indicated with a blue line. The ground truth is defined as the GPS position which is closest in time to the timestamp from when the network received measurements that initiated the location update. The GPS logging application takes 1 sample per second and is mapped to the road network (which includes footpaths, paths for cycling, and service roads), ensuring a sufficient time synchronization and accuracy between the estimated positions and their ground truth.

Table 2 summarizes the mean, standard deviation, median, and 95th percentile value of the accuracy for all scenarios (walking, cycling, and driving in urban and rural environments with a user's smartphone in standby and streaming mode).

The two basic algorithms are referred to as *cell-ID* (Section 4.1) and *centroid* (Section 4.2). The first phase of the proposed location tracking algorithm (without the route mapping filter) is referred to as *TE* (temporary estimation). The location tracking algorithm with route mapping filter, road speed limits, and mode of transportation estimation is referred to as *AMT*, named after the used inputs: antenna orientation, map, and timing information (phase II). To differentiate between the estimated locations that are available in real time and those that are

**Table 2** Accuracy, number of positioning updates, and average time and distance between two consecutive location updates, per scenario and algorithm

Scenario	Algorithm	$\mu$ [m]	$\sigma$ [m]	50th [m]	95th [m]	#Updates [-]	Update time [s]	Update distance [m]
1. Urban, on foot, 8 km, 5 km/h, standby	Cell-ID	410	287	342	1006	96	51	76
	Centroid	333	226	276	767			
	TE	270	338	150	744			
	AMT-RT	150	95	125	351			
	AMT-NTC	126	76	119	262			
	PF	165	112	139	391			
2. Urban, on foot, 8 km, 5 km/h, streaming	Cell-ID	453	324	390	1042	234	21	32
	Centroid	349	240	292	791			
	TE	205	208	151	517			
	AMT-RT	141	82	130	325			
	AMT-NTC	128	82	115	303			
	PF	158	110	137	367			
3. Urban, by bike, 8 km, 18 km/h, standby	Cell-ID	586	346	540	1127	48	32	150
	Centroid	426	270	407	923			
	TE	246	223	172	611			
	AMT-RT	189	139	158	452			
	AMT-NTC	132	83	119	296			
	PF	193	122	160	486			
4. Urban, by bike, 8 km, 18 km/h, streaming	Cell-ID	380	238	305	880	55	26	127
	Centroid	277	169	226	644			
	TE	187	254	150	384			
	AMT-RT	136	78	131	317			
	AMT-NTC	122	80	112	301			
	PF	147	103	131	331			
5. Urban, by car, 39 km, 47 km/h, standby	Cell-ID	982	587	1030	2013	58	46	600
	Centroid	808	629	661	2013			
	TE	441	486	290	1589			
	AMT-RT	370	485	243	1311			
	AMT-NTC	306	291	220	1012			
	PF	471	425	372	1322			
6. Urban, by car, 39 km, 47 km/h, streaming	Cell-ID	955	630	901	1989	91	33	411
	Centroid	780	549	645	1966			
	TE	382	398	257	1093			
	AMT-RT	336	431	241	844			
	AMT-NTC	217	141	200	467			
	PF	427	439	273	1181			
7. Rural, on foot, 8 km, 5 km/h, standby	Cell-ID	1937	1388	1764	4096	92	66	81
	Centroid	1269	983	1056	3389			
	TE	559	577	433	1469			
	AMT-RT	336	268	276	955			
	AMT-NTC	294	222	275	821			
	PF	385	313	344	1176			
8. Rural, by bike, 8 km, 22 km/h, standby	Cell-ID	2393	1310	2578	4096	37	36	196
	Centroid	1175	659	1100	2443			

**Table 2** Accuracy, number of positioning updates, and average time and distance between two consecutive location updates, per scenario and algorithm (*Continued*)

Scenario	Algorithm	$\mu$ [m]	$\sigma$ [m]	50th [m]	95th [m]	#Updates [-]	Update time [s]	Update distance [m]
9. Rural, by car, 19 km, 41 km/h, standby	TE	522	326	430	1309			
	AMT-RT	305	166	311	709			
	AMT-NTC	268	127	243	491			
	PF	391	208	389	778			
	Cell-ID	1297	1131	972	3630	43	39	380
	Centroid	746	490	670	1803			
	TE	488	594	308	1385			
	AMT-RT	280	224	188	733			
	AMT-NTC	188	186	129	589			
	PF	401	357	280	904			
10. Highway, by car, 48 km, 80 km/h, standby	Cell-ID	1059	702	1021	3096	59	35	775
	Centroid	833	537	790	1851			
	TE	352	454	231	1224			
	AMT-RT	235	232	167	826			
	AMT-NTC	138	92	122	277			
	PF	395	341	283	1253			
	Cell-ID	1045	694	984	2298			
	Centroid	700	475	612	1659			
	TE	365	386	257	1032	81	38	283
	AMT-RT	248	220	197	682			
Average over all scenarios	AMT-NTC	192	138	165	482			
	PF	313	253	251	819			

TE temporary estimation (phase I), AMT-RT real-time route mapping filter (phase II), AMT-NTC non-time-critical route mapping filter (phase II), PF particle filter

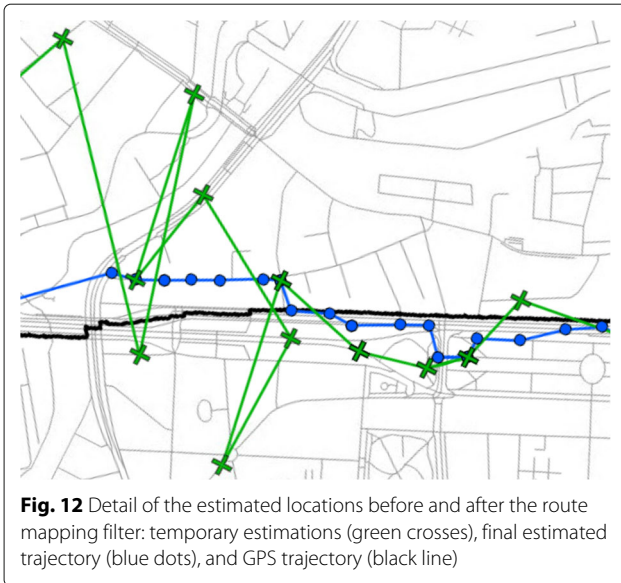
corrected by future measurements, AMT-RT (real time) and AMT-NTC (non-time critical) are used. An existing location tracking algorithm [18] based on a particle filter and map information was implemented to validate our proposed route mapping filter. These results are included in Table 2 and referred to as PF. They used regression on drive test data to estimate the probability distribution of an observation. Since drive test data is generally not available for a nationwide mobile network, the likelihood function for the particles is modified to work with the temporary estimations as input (similar to the proposed route mapping filter, ensuring a fair comparison). This particle filter is configured with 2000 particles, and the mean  $\mu$  and variance  $\sigma^2$  of the initial speed distribution are based on the mode of transportation and the maximum allowed speed of the road segments under consideration. Likewise, at each time step with measurements, our proposed route mapping filter retains the 1000 paths with the lowest associated costs in memory (MP in

Algorithm 2). The latitude and longitude coordinates from all NodeBs in the mobile network, data from the GPS logging application, and OpenStreetMap data are projected to the Belgian Lambert 72 coordinate system. Hence, the grid points and estimated locations are in the same plane coordinate reference system. This enables the use of the Euclidean distance between the estimated and actual position to define the accuracy. The total number of location updates and the average time and distance between two consecutive location updates are also included in Table 2.

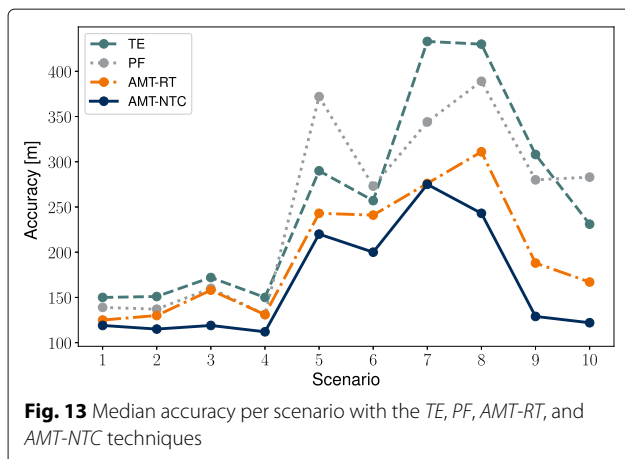
Figure 13 shows the median accuracy per scenario with the TE, PF, AMT-RT, and AMT-NTC techniques. The cell-ID and centroid approach are omitted to enhance clarity.

## 5.2 Comparison with other algorithms

It is immediately clear that the proposed location tracking algorithms outperform the classic cell-ID and centroid approach in all ten scenarios. The particle filter [18] performs slightly worse than our proposed route mapping



filter (real-time and non-time-critical version) in scenarios 1–4 and is outperformed in scenarios 5–10. The main reason for this is that the time between two location updates is variable and can be rather large (it ranges from 5 s to 6 min). In the update step of the particle filter, a new state is sampled for all particles, based on the previous state, current time, and a new random sample, and is then mapped on the road network. This can cause large deviations if the user's real speed or direction changes in this time period, which can happen multiple times during a sizeable measurement gap. The trajectories done by car and the ones in rural areas are most affected by this. In our approach, all possible locations that can be reached along the road network in this time period are considered as candidate positions for the next location update (given the previous states, i.e., user measurement history and paths in memory, estimated mode of transportation, maximum



speed limits, type of roads, and one-way street information.) The median *TE* accuracy varies between 150 and 433 m and has an average improvement, over all scenarios, of 68% and 55% compared to the *cell-ID* and *centroid* approach, respectively. The median *PF* accuracy varies between 131 and 389 m and has an average improvement, over all scenarios, of 69%, 56%, and 2% compared to the *cell-ID*, *centroid*, and *TE* approach respectively. The median *AMT-RT* accuracy varies between 125 and 311 m and has an average improvement, over all scenarios, of 74%, 64%, 20%, and 18% compared to the *cell-ID*, *centroid*, *TE*, and *PF* approach, respectively. The median *AMT-NTC* accuracy varies between 112 and 275 m and has an average improvement, over all scenarios, of 78%, 69%, 33%, 31%, and 16% compared to the *cell-ID*, *centroid*, *TE*, *PF*, and *AMT-RT* approach, respectively. The mean accuracies, standard deviations, and 95th percentile values show similar improvements.

The largest relative improvements compared to the reference algorithms are achieved with the trajectory on the highway (scenario 10). The median accuracy improves with 88% (from 1021 to 122 m) compared to the *cell-ID* approach, with 85% (from 790 to 122 m) compared to the *centroid* approach and with 57% (from 283 to 122 m) compared to the *PF* approach.

The most accurate results reported in the state-of-the-art processing techniques from Section 2 are higher than our results (accuracies up to 20 m [18], 30 m [32], and 50 m [31]), but these are achieved with synthesized data, large training sets, optimal environments, crowd-sourced measurement campaigns, and forced location update rates. However, applying the same processing technique [18] on our validation data resulted in worse accuracies but gives a realistic idea of the achievable performance without crowd-sourcing or modifications on the network or mobile side (*PF* in Table 2).

### 5.3 Non-time critical vs. real time

The non-time-critical version of the route mapping filter (*AMT-NTC*), which takes into account all measurements at once, can also work with a smaller delay (instead of at the end of a trajectory). Previously predicted locations can be corrected by multiple future measurements, but the impact tends to decrease as more time has passed between the previous update and those future measurements. For our experimental validation, this time period is 8 min; taking into account additional future measurements does not further improve the overall accuracy. Even with only 2 min of future measurement data, the mean and median overall accuracy are already 200 m and 174 m (compared to 192 m and 165 m if all future measurements are taken into account). This means that if a time delay of 2 min is allowed for the intended application, the overall mean

accuracy can already be improved by 19% compared to the real-time algorithm (*AMT-RT*).

#### 5.4 Impact of environment

The highest accuracies are achieved for the scenarios in an urban environment with trajectories on foot or by bike (scenarios 1–4). For example, the trajectory by bike in the city center of Ghent with a smartphone in streaming mode (scenario 4) has a mean, standard deviation, median, and 95th percentile value of 122 m, 80 m, 112 m, and 301 m, respectively. These higher accuracies are mainly due to the higher base station density, which is typical in urban environments. This ensures that the serving base stations have smaller separations, and hence, this limits the possible grid points because of the lower propagation delays, i.e., the green sector in Fig. 10 will cover a smaller area. When driving a car, the absolute accuracy in urban environments is worse than that in rural scenarios. For example, the improvements between two trajectories by car in an urban (scenario 5) and rural environment (scenario 9) are 63% (306 to 188 m), 56% (291 to 186 m), 71% (220 to 129 m), and 72% (1012 to 589 m), for the mean, standard deviation, median, and 95th percentile value, respectively. This is due to the sparser road network in rural areas, which increases the chance that the route mapping filter selects the correct road segments as most likely. The trajectory on the highway (scenario 10) is accurately reconstructed because the roads surrounding the highway have lower speed limits which causes these (incorrect) candidate paths to lag behind and eventually be discarded in the route mapping algorithm. Note that this is only true if there is no traffic congestion.

#### 5.5 Impact of smartphone usage

The shortest location update time or highest update rate happens when a user is walking in an urban environment while actively using his or her smartphone, i.e., through an application that sends or receives data over the mobile network on a regular basis (scenario 2). In this case, there are 234 updates during the entire trajectory, which corresponds to a location update every 21 s or every 32 m on average. Note that the update rate for this *best case scenario* is not as high as most localization algorithms for cellular networks are validated on. Location update rates of 0.5 s [29] and 10 s [18] are reported in related work, by using forced messages or synthesized validation data. Three trajectories are done for both smartphone usage modes (scenarios 1–6). The trajectories in an urban environment on foot and by bike are identical and yield similar performances for the streaming and standby mode (scenarios 1–4). The higher location update rate has a negligible impact due to the limited speed for these modes of transportation. The trajectory by car shows a significant improvement for higher location update rates (standby

vs. streaming). The accuracy increases with 41% (306 to 217 m), 106% (291 to 141 m), 10% (220 to 200 m), and 117% (1012 to 467 m), for the mean, standard deviation, median, and 95th percentile value, respectively.

#### 5.6 Impact of MoT

The trajectories done on foot and by bike yield similar accuracies as long as the environment is the same. The trajectories done by car perform worse in urban environments but better in rural environment as discussed in Section 5.4. It is to be noted that the proposed MoT estimator achieved an accuracy of 78% when a moving window of the last 5 min was used. Although this accuracy could be improved on our validation data by using a longer window, this will not always be the case, e.g., if the MoT changes during a scenario from walking to biking, a shorter window is recommended to detect the changes more quickly. Furthermore, the overall mean and median accuracy remained similar (192 m and 165 m vs. 183 m and 164 m) if the route mapping filter was provided with the correct MoT at each location update. This is because a wrong MoT estimation for a location update does not automatically result in a worse accuracy, e.g., when it is erroneously labeled as cycling while the user was actually driving at a slow speed due to traffic congestion.

## 6 Conclusions

In this paper, a technique for outdoor location tracking of all users residing on a mobile cellular network is presented. The proposed approach does not depend on prior training data and does not require any cooperation on the mobile side or changes on the network side. The topology and available measurements of a mobile cellular network are used as input for the proposed AMT algorithm (named after antenna, map, and timing information). An additional route mapping filter is applied to ensure realistic, physically possible, trajectories. The inputs for this route mapping filter are the user's measurement history, enriched open map data (road infrastructure, maximum speed limits, type of road, and one-way street information), and a mode of transportation estimator to improve the corresponding maximum speed. The novel AMT location tracking algorithm is implemented in Apache Spark to support fast cluster computing, runs completely on the network side, is confirmed to execute in real time for more than a million users in parallel, and outperforms state-of-the-art particle filters. The experimental validation is done in urban and rural environments, near Ghent and Antwerp, with experiments on foot, by bike, and by car, while a user's smartphone was used in standby and streaming mode. Improvements of up to 88%, 85%, and 57% were achieved compared to a cell-ID, a centroid, and a particle filter with map information-based location tracking technique, respectively. Future work will adapt and



apply the proposed algorithm to a 4G LTE mobile network, where further improvements are expected thanks to the more accurate timing information and the higher eNodeB density. Furthermore, the proposed algorithm will be validated on a larger test set with multiple users, different mobile devices, changes in mode of transportation, and indoor usage.

#### Abbreviations

AMT: Antenna, map, and timing information-based tracking; AoA: Angle of arrival; BLE: Bluetooth Low Energy; GPS: Global Positioning System; IMSI: International mobile subscriber identity; LBS: Location-based services; LoS: Line-of-sight; LTE: Long-term evolution; MoT: Mode of transportation; NLoS: Non-line-of-sight; PL: Path loss; RF: Radio frequency; RSCP: Received signal code power; RSSI: Received signal strength indication; TDoA: Time difference of arrival; TE: Temporary estimation; TMSI: Temporary mobile subscriber identity; ToA: Time of arrival; UMTS: Universal mobile telecommunications system

#### Funding

This research was supported by the VLAIO project ADORABLE: Anonymous Displacement and residence behavior based on Accurate mobile Location data from tElco.

#### Availability of data and materials

Data sharing is not possible for this article due to company policy of the mobile network operator.

#### Authors' contributions

JT and DP developed the novel algorithms and conducted the data analysis and interpretation. ES, MS, and MV participated in the mobile cellular data processing for the experimental validation. LM and WJ reviewed and edited the manuscript. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Author details

<sup>1</sup>Department of Information Technology, IMEC - Ghent University, Ghent Belgium. <sup>2</sup>Telenet Group, Brussels, Belgium. <sup>3</sup>RetailSonar, Ghent, Belgium.

Received: 15 November 2018 Accepted: 25 April 2019

Published online: 08 May 2019

#### References

1. F. Gustafsson, F. Gunnarsson, Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements. *IEEE Signal Proc. Mag.* **22**(4), 41–53 (2005)
2. R. Becker, R. Cáceres, K. Hanson, S. Isaacman, J. M. Loh, M. Martonosi, J. Rowland, S. Urbaneak, A. Varshavsky, C. Volinsky, Human mobility characterization from cellular network data. *Commun. ACM.* **56**(1), 74–82 (2013)
3. S. Çolak, L. P. Alexander, B. G. Alvim, S. R. Mehndiratta, M. C. González, Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities. *Transp. Res. Rec. J. Transp. Res. Board*, 126–135 (2015)
4. L. Bengtsson, X. Lu, A. Thorson, R. Garfield, J. Von Schreeb, Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. *PLoS Med.* **8**(8), 1001083 (2011)
5. OpenStreetMap contributors, *Planet dump* retrieved from <https://planet.osm.org>, (2017). <https://www.openstreetmap.org>
6. A. N. Hassan, O. Kaiwartya, A. H. Abdullah, D. K. Sheet, S. Prakash, in *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*. Geometry based inter vehicle distance estimation for instantaneous GPS failure in VANETS (ACM, 2016), p. 72
7. O. Kaiwartya, Y. Cao, J. Lloret, S. Kumar, N. Aslam, R. Kharel, A. H. Abdullah, R. R. Shah, Geometry-based localization for GPS outage in vehicular cyber physical systems. *IEEE Trans. Veh. Technol.* **67**(5), 3800–3812 (2018)
8. L. Gazzah, L. Najjar, H. Besbes, in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*. Selective hybrid RSS/AOA weighting algorithm for NLOS intra cell localization (IEEE, 2014), pp. 2546–2551
9. I. Guvenc, C.-C. Chong, A survey on TOA based wireless localization and NLOS mitigation techniques. *IEEE Commun. Surv. Tutor.* **11**(3), 107–124 (2009)
10. Y. M. Chen, C.-L. Tsai, R.-W. Fang, in *2017 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO)*. TDOA/FDOA mobile target localization and tracking with adaptive extended Kalman filter (IEEE, 2017), pp. 202–206
11. A. H. Sayed, A. Tarighat, N. Khajehnouri, Network-based wireless location: challenges faced in developing techniques for accurate wireless location information. *IEEE Signal Proc. Mag.* **22**(4), 24–40 (2005)
12. E. Xu, Z. Ding, S. Dasgupta, Source localization in wireless sensor networks from signal time-of-arrival measurements. *IEEE Trans. Signal Proc.* **59**(6), 2887–2897 (2011)
13. F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui, T. Watteyne, Understanding the limits of LoRaWAN. *IEEE Commun. Mag.* **55**(9), 34–40 (2017)
14. V. Osa, J. Matamales, J. F. Monserrat, J. López, Localization in wireless networks: the potential of triangulation techniques. *Wirel. Pers. Commun.* **68**(4), 1–14 (2013)
15. J. Borkowski, J. Lempäinen, Practical network-based techniques for mobile positioning in UMTS. *EURASIP J. Appl. Signal Proc.* **2006**, 149–149 (2006)
16. T. Wigren, Adaptive enhanced cell-id fingerprinting localization by clustering of precise position measurements. *IEEE Trans. Veh. Technol.* **56**(5), 3199–3209 (2007)
17. M. Chen, T. Sohn, D. Chmelev, D. Haehnel, J. Hightower, J. Hughes, A. LaMarca, F. Potter, I. Smith, A. Varshavsky, *Practical metropolitan-scale positioning for GSM phones. UbiComp 2006: Ubiquitous Computing. UbiComp 2006. Lecture Notes in Computer Science, vol 4206*. (Springer, Berlin, Heidelberg, 2006), pp. 225–242
18. A. Ray, S. Deb, P. Monogioudis, in *Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE International Conference On*. Localization of lte measurement records with missing information (IEEE, 2016), pp. 1–9
19. M. Ibrahim, M. Youssef, Cellsense: An accurate energy-efficient GSM positioning system. *IEEE Trans. Veh. Technol.* **61**(1), 286–296 (2012)
20. D. Plets, W. Joseph, K. Vanhecke, E. Tanghe, L. Martens, Coverage prediction and optimization algorithms for indoor environments. *EURASIP J. Wirel. Commun. Netw.* **2012**(1), 123 (2012)
21. J. Trogh, D. Plets, L. Martens, W. Joseph, Advanced real-time indoor tracking based on the viterbi algorithm and semantic data. *Int. J. Distrib. Sens. Netw.* **11**(10), 271818 (2015)
22. J. Trogh, D. Plets, A. Thielens, L. Martens, W. Joseph, Enhanced indoor location tracking through body shadowing compensation. *IEEE Sens. J.* **16**(7), 2105–2114 (2016)
23. V. Savic, H. Wymeersch, E. G. Larsson, Target tracking in confined environments with uncertain sensor positions. *IEEE Trans. Veh. Technol.* **65**(2), 870–882 (2016)
24. A. Hatami, K. Pahlavan, in *Consumer Communications and Networking Conference, 2006. CCNC 2006. 3rd IEEE*. Comparative statistical analysis of indoor positioning using empirical data and indoor radio channel models, vol. 2 (IEEE, 2006), pp. 1018–1022
25. P.-H. Tseng, K.-T. Feng, Y.-C. Lin, C.-L. Chen, Wireless location tracking algorithms for environments with insufficient signal sources. *IEEE Trans. Mob. Comput.* **8**(12), 1676–1689 (2009)
26. M. Bshara, U. Orguner, F. Gustafsson, L. Van Biesen, Robust tracking in cellular networks using HMM filters and Cell-ID measurements. *IEEE Trans. Veh. Technol.* **60**(3), 1016–1024 (2011)
27. M. McGuire, K. N. Plataniotis, A. N. Venetsanopoulos, Data fusion of power and time measurements for mobile terminal location. *IEEE Trans. Mob. Comput.* **4**(2), 142–153 (2005)
28. Y. Feng, Y. Liu, M. Batty, Modeling urban growth with GIS based cellular automata and least squares SVM rules: a case study in Qingpu-Songjiang area of Shanghai, China. *Stoch. Env. Res. Risk A.* **30**(5), 1387–1400 (2016)

29. M. Anisetti, C. A. Ardagna, V. Bellandi, E. Damiani, S. Reale, Map-based location and tracking in multipath outdoor mobile networks. *IEEE Trans. Wirel. Commun.* **10**(3), 814–824 (2011)
30. R. M. Vaghefi, R. M. Buehrer, in *Personal, Indoor, and Mobile Radio Communication (PIMRC), 2014 IEEE 25th Annual International Symposium On. Cooperative RF pattern matching positioning for LTE cellular systems* (IEEE, 2014), pp. 264–269
31. R. Margolies, R. Becker, S. Byers, S. Deb, R. Jana, S. Urbanek, C. Volinsky, in *INFOCOM 2017-IEEE Conference on Computer Communications, IEEE. Can you find me now? evaluation of network-based localization in a 4G LTE network* (IEEE, 2017), pp. 1–9
32. A. Chakraborty, L. E. Ortiz, S. R. Das, in *Computer Communications (INFOCOM), 2015 IEEE Conference On. Network-side positioning of cellular-band devices with minimal effort* (IEEE, 2015), pp. 2767–2775
33. H. Zang, J. Bolot, in *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking. Anonymization of location data does not work: a large-scale measurement study* (ACM, 2011), pp. 145–156
34. M. Arapinis, L. Mancini, E. Ritter, M. Ryan, N. Golde, K. Redon, R. Borgaonkar, in *Proceedings of the 2012 ACM Conference on Computer and Communications Security. New privacy issues in mobile telephony: fix and verification* (ACM, 2012), pp. 205–216
35. M. Arapinis, L. I. Mancini, E. Ritter, M. Ryan, in *NDSS. Privacy through pseudonymity in mobile telephony systems*, (2014)
36. A. Shaik, R. Borgaonkar, N. Asokan, V. Niemi, J.-P. Seifert, Practical attacks against privacy and availability in 4G/LTE mobile communication systems (2015). arXiv preprint arXiv:1510.07563
37. N. Husted, S. Myers, in *Proceedings of the 17th ACM Conference on Computer and Communications Security. Mobile location tracking in metro areas: malnets and others* (ACM, 2010), pp. 85–96
38. P. Newson, J. Krumm, in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. Hidden Markov map matching through noise and sparseness* (ACM, 2009), pp. 336–343
39. Propagation Delay. <http://www.telecomhall.com/analyzing-coverage-with-propagation-delay-pd-and-timing-advance-ta-gsm-wcdma-lte.aspx>. Accessed 9 Aug 2010
40. E. Trevisani, A. Vitaletti, in *Mobile Computing Systems and Applications, 2004. WMCSA 2004. Sixth IEEE Workshop On. Cell-ID location technique, limits and benefits: an experimental study* (IEEE, 2004), pp. 51–60
41. J. Wang, P. Urriza, Y. Han, D. Cabric, Weighted centroid localization algorithm: theoretical analysis and distributed implementation. *IEEE Trans. Wirel. Commun.* **10**(10), 3403–3413 (2011)

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---