Road Traffic Estimation using Cellular Network Signaling in Intelligent Transportation Systems

David Gundlegård and Johan M Karlsson

Linköping University Post Print

N.B.: When citing this work, cite the original article.

Original Publication:

Copyright: Nova Science Publishers
https://www.novapublishers.com/

Postprint available at: Linköping University Electronic Press
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-50734
Chapter 13

ROAD TRAFFIC ESTIMATION USING CELLULAR NETWORK SIGNALING IN INTELLIGENT TRANSPORTATION SYSTEMS

David Gundlegård and Johan M Karlsson
Department of Science and Technology
Linköping University, 601 74 Norrköping

ABSTRACT

In the area of Intelligent Transportation Systems the introduction of wireless communications is reshaping the information distribution concept, and is one of the most important enabling technologies. The distribution of real-time traffic information, scheduling and route-guidance information is helping the transportation management systems in their strive to optimize the system. The communication required to transfer all this information is rather expensive in terms of transmission power, use of the scarce resources of frequencies and also the building of an infrastructure to support the transceivers. By using information that already exists and is exchanged within the infrastructures of the GSM and UMTS networks, a lot of the resource problems are solved. The information that could be extracted from these cellular networks could be used to obtain accurate road traffic information to support real-time traffic information. In this way the cellular networks not only becomes the means to distribute information but also a source of road traffic information.

From the analysis made it is obvious that the potential of retrieving valuable road traffic information from cellular systems in a cost efficient way, i.e. by using already existing signaling data, is very high. It has however not been clear what to expect from these systems in terms of accuracy, availability and coverage. In this chapter the basics for this is laid out and discussed in detail. A practical trial has also been performed and the results show clearly the potential as well as the differences in using the GSM compared to the UMTS network. The advantages and drawbacks are discussed and backed up by real measurements in two different road environments. The main advantages of using the existing signaling data, i.e., passive monitoring compared to active monitoring where the terminal sends extra data is discussed and could be summarized in three components, no user acceptance is necessary, no extra signaling is necessary and it does not drain the terminal battery.
In the future it is likely that vehicles need to communicate more frequently with each other and with some kind of traffic control centre. This traffic will also be very useful in order to estimate road traffic information using the signaling information obtained from the cellular system. However, the enhanced communication systems will also change traffic patterns in the cellular networks which will affect the potential of estimating road traffic from cellular systems. The evolvement indicates that the terminals will be in active state almost constantly, and hence the updating information will be more frequent and the information more accurate.

ABBREVIATIONS

A-GPS  Assisted GPS
AOA  Angle of Arrival
BSC  Base Station Controller
BTS  Base Transceiver Station
CDMA  Carrier Division Multiple Access
CGI  Cell Global Identity
CN  Core Network
DCH  Dedicated Channel
DOA  Direction of Arrival
E-OTD  Enhanced Observed Time Difference
FACH  Forward Access CHannel
FCC  Federal Communications Commission
FCD  Floating Car Data
GPRS  General Packet Radio Service
GPS  Global Positioning System
GSM  Global System for Mobile communication
HLR  Home Location Register
ITS  Intelligent Transportation Systems
LA  Location Area
LMU  Location Measurement Units
MAHO  Mobile Assisted Handover
MEHO  Mobile Evaluated Handover
MM  Mobility Management
O/D  Origin-Destination
OTDOA  Observed Time Difference of Arrival
PCH  Paging Channel
RA  Routing Area
RNC  Radio Network Controller
RRM  Radio Resource Management
RSS  Received Signal Strength
RTT  Round Trip Time
RXLEV  Received Signal Strength
RXQUAL  Received Signal Quality
SACCH  Slow Associated Control Channel
1. INTRODUCTION

Wireless communication is one of the most important enabling technologies for Intelligent Transportation Systems. By using different kinds of wireless network technologies it is possible to distribute real-time traffic information, scheduling, route-guidance information etc. All wireless communication networks require signaling in order to support mobility management and radio resource management. This signaling information can in cellular networks, such as GSM and UMTS, also be used to estimate road traffic state information. This section describes the cellular networks potential as a source of road traffic information instead of means to distribute it.

Although the technology of using cellular networks to estimate road traffic information has been subject for analysis for quite some time, it is still far from being mature. It is not clear what to expect from these systems in terms of accuracy, availability and coverage. The potential of the system is although quite clear; it is possible to retrieve a lot of traffic data in a cost efficient way, i.e. by using existing signaling data without the need to invest time and resources in a sensor infrastructure. This is essential, especially in terms of using the existing scarce resource of frequencies and cost of building a new infrastructure.

Evolving communication systems will change communication patterns, and both the design of the communication system and the communication pattern of the users will affect the potential of estimating road traffic information from cellular systems. In order to realise a safer and more efficient transportation system, it is likely that vehicles need to communicate more frequently with each other and with some kind of traffic control centre. This traffic will also be very useful in order to estimate road traffic information in combination with the signaling information obtained from the cellular networks. Increasing the communication to and from vehicles will generate signaling traffic that in turn can be used to improve the traffic information available. Once there is an improvement in traffic information, there will be more communication to and from vehicles. Following this positive spiral, shown in Figure 1, is a potential path towards a large improvement in available real-time road traffic information.
Two fundamental characteristics are important for the development of road traffic information from cellular networks. First of all is the spatial correlation between road network and cellular network, i.e. where the vehicle tracking is challenging due to dense road network, the cellular location data has higher quality due to much smaller cells. Secondly the correlation in time between congestion and when people use their cell phones is high, i.e. when there is a lot of road traffic, there is also a lot of cellular traffic available. Outside the cities the road networks and cellular networks are sparser, see Figure 2. Furthermore, the requirements of QoS by the mobile operator are increasing, which means that the operator will have to keep track of the cell phones in more detail in the future. A drawback of this technology is that the objective of the mobile operator is to minimize the signaling traffic generated by cell phones, in order to use the network more efficient. However, this objective might shift slightly if the road traffic information becomes commercially demanded by subscribers of the operator.

Figure 1. The positive spiral of available road traffic information from cellular networks.

Figure 2. Illustration of a cellular network in Sweden, where it can be seen that the cells are small where the road network is dense and sparse outside the cities.
2. BACKGROUND

The general concept of public system architectures for mobile communication systems is a cellular network concept. The cellular networks are built based upon a number of small geographical areas, cells, covering a larger area. Each cell consists of a base station transmitting the information within the limited area. For analytical purposes these areas are usually hexagonal cells, which totally covers an area and are rather similar to real transmission patterns. Figure 3 shows a road segment and the cells covering the area as an example of how a road segment crosses several cells in its path.

![Handover point](image)

Figure 3. Travelling on a road segment involves a large number of cell border crossings. The handover between the cells are used to estimate traffic information and can be seen as virtual detectors.

The system architectures for both GSM and UMTS are rather similar in their concept. An overview of the system architectures are to be found in Figure 4. A GSM Base Transceiver Station (BTS) holds the transmit and receive equipment for one or more cells. It constitutes the interface between the network provider and the mobile phone. The Base Station Controller (BSC) administers the transmit and receive resources of the connected base stations. There are two categories of channels, signaling and traffic channels. Both the signaling channels and the traffic channel (handling the actual payload) are processed here. The concept for UMTS is rather similar, the base station is here called Node B and the device controlling a number of these cells is denoted Radio Network Controller (RNC). Figure 4 also provides the interfaces between different sections of the architecture. The interface between the BTS and BSC in GSM is denoted Abis, and the corresponding interface in UMTS is denoted Iub. The interface between the GSM systems radio parts and the core network (CN) side is denoted A and the corresponding interface for UMTS is denoted Iu.

In the core network we find the devices that connect to other networks. If the communication should go through the circuit switched network side, it uses the Mobile Switching Centre (MSC). The MSC carries out all the duties of an ordinary wireline network switch, such as processing, finding a path and supplementary services. It is also the link between the wireless networks and the wireline network. If the communication is more data oriented it would go through the packet switched network side and hence through the Serving GPRS Support Node (SGSN). In Figure 4 both devices are connected to the “network”, it should be pointed out that this is just a simplification and is in reality a number of different networks with different network features and service level agreements.
The purpose of a cellular communication system is to offer mobile communication to subscribers of the system. In order to do this the mobile operator has to keep track of where in the network a certain subscriber is located. The location is used to reach the subscriber if it has incoming data transfers and to assign the subscriber to the most appropriate radio base station. To be able to handle the mobile subscribers the operator store a number of details of each user, this is stored in the Home Location Register (HLR). Information provided by the HLR is for example if the subscriber is active or not, and if active in which part of the network the subscriber is situated at the moment.

The location within the network can with knowledge of the network structure be transformed to a position of the subscriber, i.e. coordinates. The basic operation of determining the speed of a vehicle is to measure its location two times and determine the average speed between these two points. Hence, if we could use the cellular network to determine the position of a subscriber at two times we could calculate the average speed of the subscriber between these two points. If we could do this for vehicles travelling on a road network it would be possible to estimate the average speed of vehicles on a certain road section.

![Cellular network architecture](image)

Figure 4. The cellular network architecture for the GSM (upper part) and UMTS (lower part) and their connection to the core network through the A and Iu interface, respectively.

Estimating travel times by analysing traffic from mobile phones has been subject for research since mid 1990’s. This was at a time when telematics services were supposed to expand rapidly within a few years. The cellular networks gave a possibility to communicate with the vehicles, and now it was possible to use it as a source of traffic information too. The foreseen rush of telematics services failed to come through and the willingness to pay for such services was much lower than expected. With a rather slow telematics market the interest in traffic information generated by cellular networks was reduced. One of the major problems with the telematics market was the lack of quality traffic information and this is still one of the main issues. Generating traffic information from cellular networks could be an important step in increasing the efficiency in road traffic information systems.

The task of determining a detailed position of a user in a cellular network without using dedicated positioning equipment, e.g. a GPS receiver, is challenging. Even with the most sophisticated positioning methods available today and perfect conditions the accuracy will likely not be better than 20 meters. To achieve this relatively high accuracy it is important to
keep in mind that scarce bandwidth resources are used every time a position is calculated. However, the possibilities in estimating road traffic information from cellular networks roots from the assumptions that we are only interested of positions on the road network, several measurements are available for each user and a lot of users are travelling a given road section. The first assumption makes map matching possible, which increases the accuracy significantly [1]. The second assumption gives the possibility to determine a specific route for a user. The third assumption gives statistical credibility to the data.

One of the main advantages using cellular networks for positioning is the possibility to choose the positioning points. The chosen points should be unique and stable over time in order for the algorithm to perform well. The algorithm could choose a handover point, i.e., a position where the mobile is switched from one cell to the next as it moves along a path, or a point where a certain cell has a dominating received transmission power. These positioning points are relatively easy to detect and hence a good choice to base a speed or route calculation on. However, there will not always be one of these good behaving positioning points at each end of a road section, in that case a mapping between the measured road section and the target road section is needed. The mapping could be solved by adjusting the target road sections to the available cellular data, creating new positioning points using measurement reports or simply extrapolating measured road sections to target road sections. Section 4 elaborates this issue further.

3. Cellular Network Positioning

The availability of positioning data for cellular mobile users is different depending on network, the state of the terminal, the users’ preferences and the geographical location. In this section we will elucidate some of the main properties of positioning in cellular networks. It will be based on objective facts and its merits and drawbacks will be derived on in later sections. More details in GSM and UMTS specific characteristics can be found in the referenced 3GPP standards [2-7]. For an extensive treatment of general positioning technologies and methods, see e.g. [8-10].

A standard classification of cellular positioning methods is to relate the method to which extent the network and the terminal is involved in the positioning process. The different categories are network-based, handset-based and handset-assisted.

- **Network-based positioning.** Positioning within this category rely on measurements and calculations made by nodes in the cellular network. This implies that all positioning related functionality is located in the network and that all the processing is carried out by the network. An important result of this characteristic is that the network-based positioning methods supports legacy terminals, i.e. no changes has to be made to existing terminals. Network-based positioning implies multilateral positioning, i.e. the terminal transmits data that is received by multiple network nodes. Hence, a drawback of network-based positioning is that the terminal needs to be in active mode in order for a position to be calculated. Since the cellular network determines the position of the terminal, it also falls into the category of remote positioning.
- **Handset-based positioning.** These kind of positioning methods rely on positioning measurements and calculations made by the terminal. This implies that terminals are required to be updated with positioning functionality that is not supported in legacy terminals. The main advantage with handset-based positioning is that the terminal does not need to transmit in order to calculate a position, instead unilateral positioning is used, i.e. the terminal receives signals sent by multiple network nodes and uses these for positioning. Since the terminal calculates its own position, handset-based positioning can also be referred to as a self positioning method.

- **Handset-assisted positioning.** Sometimes it is more convenient to use a unilateral approach and let the terminal report measurements to the network, where the position is calculated. If standardised measurements are used, e.g. signal strength or timing advance, legacy terminals can be used for positioning and additional functionality is not necessary in the radio access part of the network. Like network-based positioning, this is regarded as a remote positioning method.

It should be noted that remote positioning can be performed with handset-based methods by letting the terminal transmit the calculated position to the network, this is referred to as indirect remote positioning. Using the same idea, self positioning can be performed with network-based positioning, and this is referred to as indirect self positioning.

After making a distinction between where and how the measurements are performed, we will now focus on the methods used to make the positioning estimate. The more information provided the better the accuracy of the positioning estimate. The most trivial methods are based on the knowledge that the target is within a certain area. For example, once a base station register a terminal we know it is within the radio coverage of that base station (proximity sensing) or if it passes by certain radio device (signpost positioning) we know that it passed by a known position at a specific time. The signpost positioning requires a dedicated infrastructure, the denser the more accurate estimates. Another rather rough estimate is to use dead reckoning, used already by Columbus at sea. By approximating the direction and speed and apply this to the last known position we make an extrapolation which becomes the positioning estimate. This method is also known as inertial navigation. The next step is to use lateration, i.e. the range or range difference including two or more measurements. In circular lateration the distance to a reference position is known and hence we could draw a circle around this position. If we do that using one more known distance from another position we will get another circle with two intersections and hence two possible locations. The addition of a third measurement gives a unique position estimation in two dimensions. If we are only able to calculate the difference in time of arrivals of the same signal at several points we obtain a hyperbolic. The crossing of two or more hyperbolics becomes the positioning estimate (requires at least three reference stations). The advantage of using time differences is that no synchronisation between the terminal and reference station is required, on the other hand synchronisation between reference stations is required instead. These methods are also denoted Time of Arrival (TOA) and Time Difference of Arrival (TDOA), respectively.

Another method to estimate the position is to use angulation. This method is only applicable if either side is equipped with antenna arrays and hence able to detect from which approximate direction the signal is arriving. With several such sites, the terminals position estimate is restricted to a line that crosses both the target and the base station. The intersection of several of these lines becomes the positioning estimate. This method is sometimes also
denoted Angle of Arrival (AOA) and Direction of Arrival (DOA). Pattern matching, or fingerprinting, is a potential solution where some other methods fail due to shadowing or gives poor results for other reasons. The methods are based upon the fact that there exists a database with known signal characteristics from different places, usually laid out as a grid. The measured signal is compared to the database and the most equal, by some definition, entry becomes the positioning estimate.

In addition to these methods several other methods based upon range calculation by time or received signal strength (RSS) are to be found. They are, however, mostly special cases of the methods mentioned is this section. Another category of methods is hybrid approaches where the methods mentioned are used in different combinations.

There are a number of the positioning methods mentioned that are more or less standardised and in use today. For example the CGI-TA (Cell Global Identity-Timing Advance) gives information in which cell the terminal is situated and how far from the base station, i.e., a circle on which the terminal could be at any point. In the case the cell is divided into sectors we will also receive in which sector of the cell the terminal is situated. The precision of this method is coarse, however, all information already exist in the network. The corresponding method in UMTS is Cell ID-RTT (Round Trip Time). Another used method is Enhanced Observed Time Difference (E-OTD), which is based on unilateral TDOA positioning. E-OTD requires the involved three base stations to be synchronized and additional hardware in the network as well as updated cell phones. The UMTS method corresponding to E-OTD is called Observed Time Difference of Arrival (OTDOA). The multilateral versions of E-OTD and OTDOA are called Uplink TDOA (U-TDOA) in both GSM and UMTS. The TDOA-based positioning methods typically have better accuracy than e.g. CGI-TA, however, more processing, updated cell phones and new hardware is required. Also Assisted GPS (A-GPS) is standardised in GSM and UMTS for high precision positioning, these standards rely on a GPS receiver in the cell phone. More information regarding these methods can be found in e.g. [6-7].

While knowing all these methods the next questions would be the performance metrics of the position estimate. There exist several criteria for evaluating the quality of the performed estimate. Obviously, most important is accuracy and precision of the estimates. These two terms should not be mixed up or used interchangeably as often could be seen. The accuracy defines how far away from the real position the estimate is, while the precision defines how close a number of estimates is to their mean value. This means, as an example, that the precision could be high but with very low accuracy if all estimates are close to each other but far away from the real position.

An often forgotten, but very essential, performance metric category is the practical ones while comparing the different methods. The overhead of performing positioning could almost entirely be divided into signaling and computational overhead, where the former reflects the information that has to be sent over the scarce air interface resource while the latter refers to the processing power to calculate the positioning estimate. The coverage is the definition in which environments the positioning estimates should be able to be delivered, such as indoor, rural or urban environment. Some services require positioning estimates in all environments while other is restricted to specific environments. Other practical performance metrics are power consumption, especially on the terminal side which has limited battery resources, latency, i.e., the time it takes to go through the entire positioning process until an estimate is delivered. The first time this is done this is referred to as Time To First Fix (TTFF), which
could be more or less essential depending on the service. Finally, the cost also have to be taken into account, the cost could be divided into infrastructure cost to install the system and operational cost to run the system.

Positioning in cellular networks is limited to the information that is available to collect from the network. The GSM network was designed without positioning in mind, which makes the positioning related information limited. The positioning related information that is available in GSM and UMTS is described in the coming sections. A very good overview can also be found in [9].

Location Data in GSM

The backbone in an efficient travel time estimation system is analysis of signaling data generated by users in busy state, i.e. during voice calls or data sessions. This signaling data generated by busy state terminals is in GSM handled by Radio Resource Management (RRM) algorithms located in the radio access network. Complementary data can be obtained from positioning functions in the network (active monitoring) or signaling data generated by idle state terminals. Signaling data of idle state terminals is handled by Mobility Management (MM) algorithms located in the core network.

RRM is only active when the terminal is in busy state and an important task of RRM is to initiate handover. The Base Station Controller (BSC) is responsible for the handover decision and use information from measurement reports sent by the terminal and the current Base Transceiver Station (BTS). This information is very useful in the process of tracking a terminal. The terminal and the BTS repeatedly send information about received signal strength (RXLEV) and signal quality in bit error rate (RXQUAL). The fields are 6 bits long and correspond to a resolution of 64 discrete values. The terminal measures the signal quality and strength on the downlink and the BTS measures the signal quality and strength on the uplink. Based on the neighbouring list that is broadcasted by the BTS, the terminal tunes in to neighbouring cells and measures the signal strength. From the terminal, measurement reports are sent on the Slow Associated Control Channel (SACCH) once every 480 ms, the BTS adds the uplink measures and forwards a measurement result to the BSC.

Due to the propagation delay from the terminal to the BTS the terminal has to start its transmission earlier in order to avoid interference on adjacent timeslots. When the terminal shall start its transmission the delay is calculated in the BTS and the terminal is informed via a timing advance (TA) value that is sent on the SACCH to the terminal. The TA field is 6 bits long and corresponds to a resolution of 550 m. The TA value can also be used by the BSC in the handover decision and is included in the measurement report from the terminal. The BSC can use the TA value to roughly estimate the terminals velocity and, if a hierarchical cell structure is used, assign highly mobile terminals to a cell on a higher level. The TA value is also important for the BSC to complement the signal strength measurements in order to determine to which cell the terminal should be handed over.

When the terminal is in idle mode, i.e. powered on but not used for voice calls, data sessions or signaling, MM algorithms in the core network keep track of in which part of the network the terminal is located. The location information of a terminal in idle mode is much sparser and has a resolution of Location Area (LA), which consist of a configurable number of cells. The mobile terminal sends an LA update message when it detects a new LA identity.
broadcasted by the currently strongest BTS. During the LA update the terminal goes into busy state and more location information can be retrieved during a short period of time. A detailed description of GSM MM and RRM can be found in e.g. [11]. The relation between standardised location data reports and magnitudes of sampling distances are shown in Figure 5.

Figure 5. Distance between location data reports in GSM. $D_{LA}$ is the distance between location area updates, magnitude from several km up to several tens of km. $D_{HO}$ is the distance between handovers, magnitude from several hundreds of meters up to several kilometers (several tens of km in rural areas). $D_{MR}$ is the distance between measurement reports, magnitude from several meters up to several hundreds of meters.

A GPRS-attached terminal does also generate location data, the information is however slightly different from circuit switched GSM data. When the terminal is attached to the GPRS network it can be in two mobility management states, stand-by and ready state. When the terminal is in ready state it can send and receive user data. A major difference between GPRS ready state, compared to circuit switched busy state, is that the terminal itself is responsible for which BTS to communicate with (mobile evaluated handover) [5]. The terminal listens to neighbouring cells during packet transfer and decides if it should stick with the current BTS or change to a better one. In ready state cell identity, TA-value and signal strength to the serving cell is useful location parameters. Since mobile evaluated handover is used in default mode, the terminal does not report signal strength to surrounding cells in ready state and this information cannot be used for tracking the terminal. However, the network can instruct the terminal to send measurement reports if necessary. In stand-by state the terminal is connected to the GPRS network, but is unable to send user data. In stand-by state the terminal only performs Routing Area (RA) updates. A RA comprises one or more cells and is comparable to, but not the same as LA, see Figure 6 for a schematic view.

3G Differentiating Characteristics

As for GSM networks, the mobility of terminals in UMTS networks is handled by MM and RRM functions. MM and RRM are implemented in a similar manner in both systems, there are however a couple of fundamental differences. In UMTS RRM is solely handled by the UMTS Radio Access Network (UTRAN), this is achieved by connecting the Radio Network Controllers (RNC) with each other. Another important difference between the
systems is that the support for Quality of Service (QoS) for different service classes in UMTS calls for more adaptive MM. This is solved by implementing MM functions not only in the core network, but also in UTRAN. More information regarding UMTS MM and RRM can be found in e.g. [12-13].

In both GSM and UMTS the MM state of the terminal decides how much location information that is available. The MM state model in UMTS reminds a lot of the one used in GSM/GPRS, although the UTRAN MM adds a number of new states. Principally the location of the terminal in UMTS is known on cell level and mobile assisted (network evaluated) handover (MAHO) is used when the terminal is used for a service with high QoS demands (e.g. speech or a high bit rate data session). When the terminal is switched on but not used for data transfer it is known on LA or RA level. When the terminal is used for low bit rate data transfer it is known on cell or UTRAN Registration Area (URA) level depending on mobility.

![Relation between Location Area (LA), Routing Area (RA), UTRAN Registration Area (URA) and cell.](image)

The location of the terminal is known in most detail when the terminal is used for circuit switched services or high speed data services, i.e. when the UTRAN MM state is Cell DCH (Dedicated Channel). In this state MAHO is used, which means that the terminal continuously reports data of the radio connection that can be used to locate the terminal. This state is similar to busy state of circuit switched GSM. The RRM of the two systems are however quite different, which leads to a number of important differences in available information. The main differentiating characteristics of UMTS RRM compared to GSM are:

- Handover control
- Time alignment
- Power control
These functions will affect the available information of a connection. The most important difference in handover control is the use of soft handover in UMTS. This means that the terminal is connected to several base stations at the same time and the location of the terminal can be determined in more detail. The terminal is expected to be in soft handover during 20-40% of the time [13]. Another difference in handover control is how the network makes the handover decision, i.e. the characteristics of the measurement reports that are sent from the terminal and the base station to the RNC.

In UMTS the periodic measurement report interval is configurable between 0.25 and 64 seconds, depending on radio environment and the state of the mobile terminal [14]. The frequency of event triggered reports are dependent of the frequency of actual events, e.g. a new radio link addition to the active set, but also on the operator configurable parameters time-to-trigger, hysteresis and offset value. More detailed information on UMTS measurement reports can be found in e.g. [3-4, 14]. Signal strength and quality of serving base station(s) are similar to the ones in GSM. The maximum number of surrounding base stations that can be measured is increased from 6 in GSM to 32 in UMTS. A TA value is not calculated in UMTS (WCDMA) networks since it is not a TDMA based system, but other time alignment measurements are available, e.g. round trip time and time difference between base stations [4].

The soft handover used in UMTS means that a terminal can be connected to several base stations simultaneously, whereas in GSM the terminal is only connected to one base station at the time. To track a vehicle, both measurement reports containing radio parameters and handover points can be used. When it comes to calculating travel times it is important to have two accurate estimations of the vehicle’s position in order to make a good estimation of the travel time between those points. The handover points in GSM are a good candidate to estimate those positions. However, in UMTS the terminal will not change from one base station to another, instead radio links will be added to and removed from the terminals active set. As shown in [15], soft handover points in UMTS seem to be very promising to use for travel time estimation.

A potential problem with travel time estimation is the use of cell breathing, which probably will be more utilised in UMTS than in GSM. In cell breathing the size of the cells can change dynamically depending on the capacity need of different areas. An important tool to determine travel time is to measure the time between handover points, and if cell breathing is used the handover points will change with time. It will hence be more difficult to predict the handover points, but the pilot power of the base stations is known and can be used as input for the predictions.

An important function of the Code Division Multiple Access (CDMA) based systems is fast power control. In UMTS the inner loop power control makes it possible to adjust the terminal’s power level 1500 times per second [13]. This can be compared to approximately two times per second in GSM. This means that the power level between the terminal and its serving base station(s) is measured at least 1500 times per second, and hence a massive collection of data is available. This data can be used to locate the terminal relative the connected base station(s) and be useful in the travel time estimation. However, a potential problem is that the inner loop power control is performed in the base station, which means that the information is not available in the RNC where it is viable to collect it. The RNC is though responsible for the outer loop power control, i.e. to control the target signal to interference ratio (SIR), which has a frequency of 10-100 Hz [13].
Reports to support handover and power control will be useful in order to locate the terminal according to relative received power level from different base stations. Power levels might not be the most efficient measurement to use when a terminal shall be located; this is due to fast and shadow fading. More often time differences are used to calculate the position of a terminal, which leads us to the time alignment comparison of UMTS and GSM. As described, time alignment is important in GSM and is managed with the TA-value calculated by the BTS and sent to the terminal. Since UMTS is a CDMA based system, the time alignment is not needed in order to avoid co-channel interference and is not implemented. However, similar time-alignment measurements are also available in UMTS.

During soft handover it is important to minimize the buffer needed in the terminal to combine the signal from the base stations. To do this, the terminal measure the time difference between the base stations and send this information to the network, which compensates for this by time alignment of the base station signals. The time difference is measured in terms of time difference between the system frame number (SFN) of different cells and is often referred to as the SFN-SFN time difference. If we know the real-time difference between the base stations, which can be measured by base stations or location measurement units (LMUs), we can narrow down the position of the terminal relative the base stations (cf. TDOA positioning). Another possibility is to use the round-trip time (RTT) measurements calculated by UTRAN. Reference [4] states that the RTT should have a measurement period of 100 ms and an accuracy of $\pm 0.5$ chips. One chip accuracy in time correspond to approximately 80 m. Reference [16] claims however that it is possible to measure RTT with the accuracy of $\frac{1}{16}$ of a chip, which corresponds to approximately 5 m.

So far terminals in Cell DCH state have been discussed. However, terminals in Cell Forward Access Channel (FACH) and Cell Paging Channel (PCH) state will be very useful in travel time estimations since the terminal performs cell updates in both of these states. Depending on the size configuration of URA, also terminals in URA PCH state might produce useful information. Terminals in these three states are characterised by having small amounts of data with low QoS demands to transfer [12]. If the URA is configured to be large, the information from URA updates can be used for the same purpose as RA or LA updates, e.g. as input for O-D matrix estimations or traffic flow measurements over large areas.

Another fundamental difference between UMTS and GSM is the physical layer implementation. The characteristics of the modulation and wide spectrum of UMTS make it more suitable for positioning [8]. These characteristics may also affect the possibility to estimate the speed of the terminal according to the reception properties of the signal. This type of speed estimation can be useful in estimating travel times; it will however depend on implementation of the technique in the cellular networks. Different solutions to do this are described in [17-19].

The physical layer implementation also calls for another crucial characteristic of UMTS compared to GSM, in general significantly smaller cells. A denser network of base stations makes the location accuracy better, both in terms of active and passive monitoring as described in the following sections.
Table 1. Location data relevant for road traffic information estimation in GSM and UMTS networks

<table>
<thead>
<tr>
<th></th>
<th>GSM</th>
<th>UMTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synchronization level</td>
<td>1108 m</td>
<td>78 m</td>
</tr>
<tr>
<td>[bit length in m]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time alignment</td>
<td>TA</td>
<td>RTT/SFN-SFN</td>
</tr>
<tr>
<td>Cell size</td>
<td>&lt; 35 km</td>
<td>&lt; 10 km</td>
</tr>
<tr>
<td>Registration areas</td>
<td>LA/RA</td>
<td>LA/RA/URA</td>
</tr>
<tr>
<td>Measurement report interval</td>
<td>Periodic 0.48 s</td>
<td>Periodic 0.25-64 s</td>
</tr>
<tr>
<td></td>
<td>Event-triggered</td>
<td></td>
</tr>
<tr>
<td>Max number of cell measurements</td>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td>Power control frequency</td>
<td>2.1 Hz</td>
<td>Inner loop: 1500 Hz</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Outer loop: 10-100 Hz</td>
</tr>
</tbody>
</table>

4. ESTIMATION APPROACHES

Commercial companies have contributed to a large part of the development in estimating road traffic information from cellular networks in recent years. This means that few research publications are available describing the estimation approaches. Good exceptions are the work of e.g. Zhi-Jun Qiu et al. at University of Madison Wisconsin, see e.g. [20-24], and Bruce Hellinga et al. at University of Waterloo, see e.g. [25-27]. In order to estimate road traffic information from cellular networks the following basic steps can be included:

1) Location data collection
2) Terminal classification
3) Map matching
4) Route determination
5) Traffic state calculation

The location data collection phase involves how to gather the relevant data from the cellular network. The available data is described in section 3. The terminal classification focus on determining which of the mobile terminals that are located on the road network and in which type of transport mode they are. The map matching phase includes how to connect the location data to the road network, whereas the route determination process is used to determine the route of the vehicle, consisting of several road segments. The last step is to perform the traffic state calculation. Depending on the traffic information that is of interest, e.g. long term traffic flow estimation, origin-destination (O/D) estimation, travel time estimation and incident detection, different approaches and location data will be used. The steps can be carried out in a different order, in combination with each other and there can also be iteration between the different steps.
Location Data Collection

Two different approaches can be distinguished when collecting location data from the cellular networks. The first approach is based on the Federal Communications Commission (FCC) mandate that all mobile phones should be possible to locate with certain accuracy. A similar agreement exists in the EU countries and the operators active in EU. This implies that mobile phones can be located periodically and hence an average speed can be determined for the vehicle. This approach has the drawback that it generates extra traffic in the network and might be more vulnerable to privacy issues. The second approach relies on monitoring the generated traffic without trying to explicitly locate any of the mobile phones. The first approach is referred to as active monitoring whereas the second is referred to as passive monitoring.

Active monitoring can use any of the standardized positioning technologies described in section 3, e.g. CGI-TA or E-OTD. In passive monitoring it is possible to use all the data generated by the terminal to track a vehicle, this data is also described in detail in section 3. The problem with the passive monitoring approach is that terminals only generate detailed location information in busy state. This reduces the number of available probes significantly compared to all mobile phones that are switched on. The hybrid approach is based on passive monitoring with the possibility to complement with active monitoring upon certain criteria, e.g. large variations in estimations and light network load.

The possibility of estimating road traffic information with the help of cellular networks is well known, although the technique is not widely used today. Several commercial solutions are available, most of them situated in the U.S. Commercial companies, public organizations and universities have carried out field tests and simulations in order to evaluate the possibilities of the technology. A number of tests have been carried out in the U.S., but field tests have also been carried out in for example Austria, Canada, China, Finland, France, Germany, Israel, the Netherlands and Spain, see e.g. [28-35].

The Cellular Applied to ITS tracking And Location (CAPITAL) project was one of the first attempts to exploit cellular data to extract traffic information. The operational project started in 1994 in Virginia and ran for 27 months. The system used TOA together with AOA positioning to actively monitor different subscribers. The solution is based on active monitoring and it was unable to extract any useful information [28, 36]. Since then a lot of experience has been made in field tests and simulations and a number of projects have reported promising results.

The projects following CAPITAL have taken many different approaches to extract information from the cellular networks. Early papers in the area are mainly focused on the active monitoring approach. Also the developed simulation models are based on this approach. Reference [37-39] use simulation to evaluate the impact of system parameters, e.g. sampling interval and positioning error, in active monitoring based systems. Also [40] is an active monitoring based simulation model, and the focus is to evaluate a segment based approach to estimate travel times.

A number of papers assume configurations to the cellular network in order to generate more detailed signaling traffic [41-45]. These systems will be able to estimate the traffic conditions better than standard passive monitoring systems, but it is doubtful that the

---

1 AirSage, Applied Generics, CellInt, IntelliOne, ITIS Holdings, TrafficCast.
signaling configurations will be implemented in a commercial communication system. A couple of field tests has been carried out where cell phones are altered in order to send location and speed data with regular intervals, it is important to distinguish these tests since they require user acceptance and special software and is better categorized in the area of regular floating car data (FCD). It is possible that also active monitoring will require user acceptance since arbitrary switched on cell phones can be tracked and the tracking actually drains the terminal battery.

Due to an increasing number of mobile terminals on the roads that generate useful data, a wish to minimize network load and better tracking algorithms, the passive monitoring approach has gained popularity. Several ways of passively collecting signaling data from the network are proposed. The first one is based on analysis of billing information sent from the core network. This approach is used in [31, 46] and makes it easy to collect the data typically in a single interface for the whole network. The billing information is not as detailed as the information available in other parts of the network and therefore systems are proposed using either the A- or Abis-interface of the GSM network. If the A-interface is used, fewer installations have to be made for a certain geographic area. On the other hand the available location data is not as detailed as in the Abis-interface. Basically the difference is that in the A-interface only handover and location area updates can be used, whereas in the Abis-interface also measurement reports can be used to estimate the location of the terminal. Passive monitoring via A- or Abis-interface is used in e.g. [32, 47, 48].

The natural extension to use either active or passive monitoring is to combine these two approaches; this is also suggested in recent commercial systems [49-51]. It is, however, unclear from present publications if it has been evaluated in a field test. This approach makes it possible to gather more information when it is most useful without putting unnecessary load on the network.

**Terminal classification**

Large amounts of signaling data is generated in a cellular network. However, we are only interested in tracking the terminals that are travelling on the road network. Furthermore the objective is often to estimate the travel time as it will be experienced by as many users as possible. This means that we are not interested in tracking motorcycles, bicycles, buses and taxis in separate bus lanes etc. since their travel times are not applicable to most of the road users. Another problem is that multiple cell phones in a single vehicle or bus will bias the travel time estimation.

First of all there is a need to filter out users that are not related to the road network, e.g. stationary users and pedestrians. This step can be performed in combination with the traffic state calculation. Stationary users can be filtered out by ignoring terminals that has a speed, as determined by e.g. handovers, below for example 6 km/h [50]. This means that terminals not moving fast enough will never be considered in the traffic state estimation. A similar approach is described in [22], with the difference that outliers are filtered in relation to the average speed in the previous time period. An interesting problem arises during congestion when vehicles are travelling at a lower speed than the chosen threshold. However, a straightforward way of solving this is to consider terminals that recently have been registered for a higher speed as valid probes, although they currently might be travelling at a lower
speed than the threshold. Using this approach we will sometimes classify valid terminals as pedestrians during severe congestions. However, these severe congestions are most likely best detected using dedicated algorithms and possibly using other type of data.

As mentioned, buses and taxis can bias the estimations if they travel in a separate lane. Also motorcycles, bicycles and even trains can be a problem in some cases. Taxis, motorcycles and bicycles are very difficult to detect unless we have many samples in a time interval. In [52] a promising method to classify buses and possibly also trains using their timetables is described.

**Map Matching and Route Determination**

Since the location data from the cellular network is relatively coarse, the map matching step is quite important. Map matching and route determination is often combined. Depending on whether active or passive monitoring is used the map matching approach will differ. If active monitoring is used, due to the relatively slow sampling rate, an approach similar to map matching for regular FCD is used, i.e. variants of point-to-curve matching. Since the location accuracy is quite low, using previous locations and connections between different sections is more important compared to GPS-based systems. Map matching for active monitoring systems is described in detail in e.g. [27, 29]. If passive monitoring is used the sampling rate (from measurement reports) will typically be much higher, which means that some kind of curve-to-curve matching might be interesting. A difference with passive monitoring is also that the terminals are located spatially uniform, i.e. at the same place every time a travel time is calculated, which means that the map matching task is more predictable.

If the task of map matching involves how to match one or more location samples to a road section, the route detection task is to determine the order of sections that has been traversed. This can be done using the terminal locations in combination with the network structure and section decisions. Useful input to the route determination is also the measurement reports, since pattern matching techniques can be used to determine the most likely section or route. The tracking of cell phones using measurement reports in combination with Markov Models and Kalman filtering are discussed further in [53, 54].

When determining the route and matching positioning measurements to points on the road, a radio map including handover locations and signal strengths is needed. This map can be estimated using propagation models, test drives or a combination. Although test drives require more work, most likely they will also improve the accuracy of the radio maps and hence improve travel time estimations.

**Traffic State Calculation**

Once the section and/or route has been determined, the traffic information can be calculated. For O/D-matrix and traffic flow estimation the route determination is unnecessary, in that case map matching is the last step before state estimation. Handovers and measurement reports are not really useful in O/D-estimations since the number of active terminals is so few. However, location area and periodic updates are very useful since they are performed by all cell phones in a network. As described in section 3, all cell phones report
their location when a telephone call is in progress, when an SMS is sent, when it is used for data transfer, when the location area is changed and also periodically (e.g. every fourth hour). This information is very useful in the O/D-matrix estimation process. If a cell phone is tracked during a longer period of time (at least several days) a good estimation of the travel behaviour can be made. Details of O/D-matrix and traffic flow estimation can be found in e.g. [55, 56].

In order to estimate travel times, two accurate locations with suitable separation are needed. These locations are straightforward using active monitoring, since it is the periodically collected position estimates. In passive monitoring a lot more information is typically available from the active terminals, and we can choose to estimate the position of the terminals at locations that are suitable from both a travel time and positioning accuracy point of view. As illustrated in Figure 7, the location can be estimated using handovers or by analysing measurement reports and defining proprietary location triggers. The potential gain in using proprietary triggers is that the handovers are not optimised to estimate positions, but instead optimising cellular network performance. The handovers can for example be a function of network load or interference, which can give a bias to the predicted handover location. The drawback of using a proprietary location trigger is that a lot more processing is needed.

\[ \text{SNR} = \text{Signal-to-Noise Ratio} \quad t = \text{time} \quad L = \text{Network Load} \]
\[ \text{SIR} = \text{Signal-to-Interference Ratio} \quad P = \text{max transmitted BTS power} \]

Figure 7. Different approaches to position estimation using measurement reports for travel time estimation.

Whether active or passive monitoring is used, the travel time between estimated locations needs to be mapped to the target links, where travel times are supposed to be reported to travellers. In active monitoring it is straightforward to give weights to the measurement in
relation to the distance travelled on the target link [57]. As mentioned previously, in passive monitoring it could be solved by adjusting the target road sections to the available cellular data, creating new positioning points using measurement reports or simply extrapolating measured road sections to target road sections. Another possibility is that a target road section consists of several estimated positions, and in this case the estimated travel times should be weighted in relation to their length compared to the total length of the section.

There are only a limited number of traffic state estimation models for passive monitoring available in the literature. The traffic state is in [21] estimated using handovers and both a first-order model including average speed and a second-order model also incorporating estimated traffic flow. The estimated state is then filtered using a particle filter. In [22] handovers and a first-order model in combination with a Kalman filter are used to estimate average speed.

Incident detection can be carried out using travel times, traffic flow, or a combination of the two estimations. However, as first described in [37], there is also a strong correlation between congestion and the number of phone calls, since people tend to start making phone calls when entering a queue.

5. NETWORK CHARACTERISTICS AND DATA QUALITY

The quality of the road traffic information estimated from cellular networks is dependent on a number of aspects related to the geographic location of the implementation. The potential quality depends on the cellular network structure and configuration as well as the road network structure together with traffic characteristics of both the cellular and road network. The actual performance will also be affected by the implementation of the positioning, tracking and filtering methods.

The structure of the cellular network determines the available data for position estimation. Smaller cells give generally less ambiguity in position estimations and also more handover locations that can be used for travel time estimation. The route determination is also easier to perform with more cells within range. A configuration made by the mobile operator is how to design the registration areas, i.e. how to group the cells into URA, RA and LA, and this will also affect the available location data.

Correspondingly, the structure of the road network will also affect the challenges in extracting useful information. One of the major challenges today is to estimate high quality travel times using cellular networks in urban environments. The challenge using this kind of system in an urban environment lies mainly in the large number of non-vehicle terminals and possible routes. Furthermore, the mean speed of vehicle terminals is not much higher than non-vehicle terminals, especially in the case of congestion. The travel times are also exposed to large variations due to waiting times in signalized intersections. An interesting approach to overcome the problems of large variations in urban environments is described in [25].

The traffic characteristics of the cellular network are naturally important for the estimation quality. The road traffic characteristics are affecting indirectly, i.e. few cars means few cell phones and hence few measurements. This is not a large issue since it is most often no traffic related problems when the traffic flow is low. The number of calls in the cellular network is however a key indicator of the quality. Since the measurements are relatively
noisy, it is important to have several samples in a time period in order to achieve high quality estimations. The relation between number of samples and estimation quality is discussed in detail in [24]. Also the lengths of the phone calls are of great importance, since the cell phone needs to pass several location points (e.g. handovers) in order for a travel time to be estimated.

The traffic information quality is also to a large extent affected by the location accuracy, which is discussed further in the remaining part of this section.

Location Experiments

As described earlier, travel time estimation require at least two reasonably accurate location estimates. The accuracy of these locations also determines the accuracy of the travel times. In order to estimate these locations, measurements of channel characteristics such as signal strength or propagation time are needed. These measurements can be obtained from the access network during terminal transmission, from measurement reports made by the mobile terminal or both. The time interval between measurement reports is crucial for the positioning accuracy and can be used to define a lower bound on accuracy. Whether the terminal location is estimated based on handovers or on raw measurement reports, the event is based on the measurement reports. The potential error introduced by the measurement report interval corresponds to half of the actual measurement report period. In GSM the measurement report interval is 480 ms, which corresponds to a location error of 2 meters driving at 30 km/h and 8 meters driving at 120 km/h. As described previously, the measurement reports in UMTS can be event-triggered or periodic. If the measurement reports are event-triggered, there is theoretically no error addition. However, the measurement report will most likely be delayed to an integer number of the shortest periodic reporting interval. The shortest periodic time interval is 250 ms and this corresponds to 1.0 and 4.3 meters for 30 and 120 km/h, respectively. In UMTS the periodic interval is configurable by the operator between 0.25 and 64 seconds, which means that the quality will depend on operator configuration. The error introduced by the measurement report seems in most cases negligible, hence other factors such as signal variations due to e.g. multipath propagation and handover inconsistency due to interference contributes to a larger extent to the location error. To get an idea of the potential of road traffic estimations from cellular networks an empirical study of the handover location error in GSM and UMTS is performed. For a more detailed discussion of the results, see [15].

The aim of the location experiments is to compare the location accuracy of handover points in GSM and UMTS. The first experiments were carried out on a 900 meters long street segment in a “sparse” urban environment. The segment was driven fifteen times back and forth with test equipment for GSM and UMTS (Ericsson TEMS Investigation 7.0) and a GPS receiver. Signaling data was collected from a GSM terminal and a UMTS terminal simultaneously, both with ongoing telephone calls. For comparison another test run was made in a different environment. The second test was performed in a suburban highway environment with less complex GSM cellular structure. The handover locations of the sparse urban environment are shown in Figure 8.
Figure 8 indicates quite large differences in the behaviour of the GSM and UMTS networks regarding handovers. In the suburban highway environment several UMTS and GSM handover zones were detected. Three of these handover zones for both GSM and UMTS were analysed in detail using seven test rounds on a 1.5 km road segment.

The percentage of times a specific handover is completed in the same handover zone, driving the same route several times, is referred to as consistency. The consistency of handover points will affect the travel time estimations. If a predicted handover does not occur, it will feed the system with no data or corrupt data, which will affect also an averaged calculated travel time. The consistency of the different handover points for GSM and UMTS are shown in Figure 9.

![Handover locations in UMTS (left) and GSM (right).](image)

**Figure 9.** Sparse urban environment. Handover consistency in UMTS and GSM.
Compared to the UMTS case, the GSM handovers were much more scattered. Several handovers occurred outside the handover zones and the handovers within the zones were between different cells. The reason that two of the UMTS zones did not have 100% consistency was that no soft handover occurred in these zones in several of the rounds. The GSM inconsistency on the other hand, was due to handover between different cells in the same handover point. An explanation of the behavior of the UMTS and GSM networks can be that the UMTS network has more free capacity than the GSM network, and there is a possibility that the inconsistency of the GSM handovers are due to cell capacity limitations. However, the test runs were driven at two different days and times and it is not likely that the network was congested at both occasions. Since the results were similar for GSM in both test runs, a congested network does not seem to be a good explanation. A more plausible explanation might be the different cell structures of GSM and UMTS in the test area. Due to relatively few subscribers in the UMTS network, a hierarchical cell structure has not yet been deployed, as often is the case for the GSM network. A hierarchical cell structure might explain some of the scattered handovers and the fact that several different handovers were detected in the same handover zone.

The handover location accuracy is assessed using data from the 15 rounds to calculate an average handover point and measure the deviation from this point in the different rounds. The deviation is measured from the average point of all handovers within a handover zone. It should be noted that since the GSM handovers were very inconsistent, handovers between different cells will be grouped to the same average point, which will affect the GSM handover location accuracy measurements negatively. The handovers in UMTS are defined as the first or last radio link addition depending on handover zone. Missing handovers in a zone are ignored in the accuracy measurements, i.e. the consistency and accuracy diagrams should be assessed and analysed independently. All handover positions are mapped to the one-dimensional space of the traversed road segment. The positioning error of the GPS receiver is included in all measurements.

---

**Figure 10.** Sparse urban environment. Mean and max handover location error (in deviation from average handover point) together with the standard deviation of the measurements.
Figure 10 shows that the accuracy is quite good for a majority of the handover zones in both GSM and UMTS. The mean error is below 20 meters for more or less all of the UMTS handover zones, and below 40 meters for the GSM zones. However, a large variation in accuracy between different handover zones can also be seen, as well as a large difference between mean and max error. The large max error of handover zone 4 in UMTS was due to an outlier in one of the rounds. The UMTS accuracy seems in general much better than for GSM, which, as mentioned previously, partly could be explained by different cell structure and traffic pattern in the area. It should be noted that it is possible that the soft handover technique used in UMTS might render much more stable handover points compared to GSM, but it is not clear from the experiments what will happen if the traffic pattern and cell structure of the UMTS network changes. In the measurements the true value of the average handover point for the data set is used to calculate the location error. This value will not be known in a travel time estimation system, but instead estimated using test runs or coverage maps.

The previous hypothesis that the low accuracy and consistency of the GSM handovers is due to a complex hierarchical cell structure is evaluated with a second test in a different road and radio environment. This environment has most likely a flat GSM cell structure and the road is a highway with speed limit 90 km/h. In this environment it is also possible to evaluate how the vehicle speed affects the handover accuracy. It is possible that a higher speed gives a larger handover zone due to larger movement between measurement reports. The results of the test runs are shown in Figure 11.

The consistency for all evaluated handover zones is 100% for both GSM and UMTS. This implies that the low consistency for GSM in the urban environment can be due to the cell structure. The location accuracy of the UMTS handovers is in the same order as the best ones from the urban environment, i.e. very good. On the other hand, the GSM measurements have a large variation where one handover zone has an average error of 20 meters where as another one has an average error of almost 80 meters. From the results we can expect that the vehicle speed does not have a dramatic effect on the accuracy, the handover consistency can
be largely affected by cell structure and the UMTS handover location accuracy seems to have much higher location accuracy.

*Figures 12 and 13* show the actual travel time and the estimated travel time together with actual mean speed and estimated mean speed for the street segment traversed in the test runs. The travel times and mean speeds are calculated between the average handover points in the GSM network (handover zones 3 and 4) and the UMTS network (handover zones 3 and 4). The actual travel time is calculated from the TEMS log file between the average handover points for all rounds. The estimated travel time is calculated from the TEMS log file between the actual handover points in the current round.

It is important to mention that the travel times are calculated under the assumption that all vehicles are tracked to the correct road segment and that it is identified as a vehicle. In a real system, these tasks are typically very challenging, especially in an urban environment.

![Travel Time and Mean speed - GSM North Direction](image1)

Figure 12. Travel times and mean speed estimated with GSM handover zones 3 and 4.

![Travel Time and Mean speed - UMTS North Direction](image2)

Figure 13. Travel times and mean speed estimated with UMTS handover zones 3 and 4.
As can be seen in Figure 13, the estimated UMTS travel times follow the actual travel times extremely well. However, as can be seen in Figure 12, also the estimated GSM travel times follow the actual travel times in a good way. Four main factors affecting travel time accuracy are handover location accuracy, handover consistency, number of available probes in a time interval and road segment length. Handover location accuracy and consistency have been discussed, but also the number of available probes and the road segment length are important for travel time accuracy, interestingly they are also correlated. Increasing the road segment length will make the handover location error less significant and yield better travel time accuracy. More vehicle probes will give a better accuracy due to the possibility of averaging the travel time value between several probes in a travel time reporting interval. However, increasing the road segment length will also decrease the number of ongoing telephone calls or data sessions that completes the whole segment, which gives less probes for averaging. Shorter road segments for travel times also lead to more detailed traffic information, which might be useful in for example urban environments.

The road segment lengths (distance between average handover points) of the experiments were 750 and 785 meters for GSM and UMTS, respectively. The experiments indicate that it is possible to make good travel time estimations from cellular networks for this segment length, even shorter segment lengths are possible depending on radio environment, number of probes and accuracy requirements. Another factor that will affect the travel time accuracy is the average vehicle speed in the handover zones. If a vehicle is travelling slowly in a handover zone, a small location error can affect the travel time quite drastically. This problem will typically be common when deploying this kind of systems in urban environments.

6. SENSOR COMPARISON AND MEASUREMENT OBJECTIVES

Numerous sensor types are available to measure road traffic state information such as speed, density and flow. Stationary sensors, e.g. inductive loops and IR sensors, measure vehicle traffic parameters in a given location. Floating sensors are located in vehicles and measure the parameters for a given vehicle at different locations. The vehicles that are equipped with sensors are often referred to as probe vehicles or floating cars. License plate matching technologies measure the travel time between video camera locations. Different types of sensors have different advantages and drawbacks. Which sensor that should be used is dependent on traffic conditions, road network structure and financial aspects, but also the main application of data is relevant for this choice.

The performance of different systems for traffic information depends on a number of factors related to the measurement procedure and the number of sensors, but also on how the performance metrics are defined. Ideally we would want to know our own travel time given that we start our travel at a certain time in the near future. However, that is a quite challenging task, and the aim is currently to report the historical travel time dating e.g. five minutes back. Travel times fluctuate due to individual driving patterns and it is not obvious if we want to know the lowest, average or the highest travel time in the reporting interval. Speed fluctuations due to driver behaviour decreases when the road gets more congested and the incident or slowdown detection is another important application. However, even in a congested situation the travel time between different lanes can differ significantly. Figure 14
shows a good example of the variability in travel times measured by GPS equipped vehicles travelling the same route simultaneously. The test was performed by ITIS Traffic Services Ltd on Ayalon freeway in Tel-Aviv.

![Image](image.png)

**Figure 14.** Variability of individual travel times travelling the same multi-lane road section simultaneously (Source: ITIS Traffic Services Ltd).

A common performance metric is the time for a system to detect an incident. The definition of an incident varies, e.g. 40% slowdown, and affects the performance. For stationary sensors the time to detect an incident depends on the time it takes for the incident to propagate to the sensor. The time for the incident to propagate to the sensor depends on the average speed of the vehicle, which typically is very low during an incident. In this case the sensor spacing is the crucial factor and a dense network of sensors is related to large investment and maintenance costs. For floating sensors the time to detect an incident depends on the time for the sensors to propagate to the incident. This time is highly correlated to the number of vehicles with sensors (penetration) and the reporting interval of the sensors. Using signaling data from cellular networks gives a potentially high number of floating sensors with a very short reporting interval. The issue in these kinds of systems is the relatively low location accuracy discussed in previous sections. Since the time to detect incidents is dependent on sensor spacing for stationary sensors and sampling interval for floating sensors, ideally we would like to increase the number of sensors and decrease the sampling interval in congested areas. For stationary sensors this is rather difficult. For regular FCD systems and active monitoring systems this is achieved by the increase in number of vehicles in congested areas, and it is also quite easy to modify sampling intervals of these probes depending on the
traffic conditions. Passive monitoring systems also benefit from the increased number of vehicles, but also from the fact that cell phones are more often used during congestion. As previously described, hybrid monitoring systems can combine the advantages of passive and active monitoring systems.

Another important performance metric is the travel time or average speed deviation relative some kind of ground truth. As previously explained and as can be seen in figure 14, it is not straightforward to determine the ground truth. Travel times from cellular networks are often evaluated against loop detectors, which measure the average speed at certain locations and can only be used to estimate the travel time. It is also quite common to use GPS equipped vehicles to measure experienced travel times for individual cars, which might vary significantly between different drivers and lanes. Although it might be costly, a good way of evaluating travel time accuracy is to use license plate matching cameras.

In the comparison between different sensors, it should be noted that all floating sensors rely on stochastic elements, i.e. that floating cars are travelling in all areas with a minimum frequency and that cell phones are used in vehicles. Leaving out effects of extreme weather conditions and severe congestions, stationary sensors are more deterministic in nature and have a vehicle penetration of almost 100%. Considering this, the best results are most likely obtained when information is merged from several sensors with different characteristics.

7. CONCLUSION

Since the failure of the pioneering CAPITAL project in 1994 the estimation technologies have been improved tremendously. A couple of years ago successful field trials started showing up in different parts of the world with different technology providers and since then the number of field trials and large scale implementations have increased dramatically. Recently three extensive reports have been published summarising the different projects and field trials that has been performed within the field of estimating road traffic information from cellular networks [31-33]. The conclusions from these reports are quite similar; the technology has large potential although more research is needed in order to use it as efficiently as possible. The technology has shown very good results on highways during light and medium congested traffic. However, during severe congestion the results are varying between different vendors and different trial areas. Promising results have been shown in arterials and even urban environments, the results in these environments are though quite few and very varying. Problems have also been identified during the morning rush, since people are reluctant to use the cell phone early in the morning.

Detailed analysis of specific deployments of commercial systems have been made recently in e.g. [58, 59]. However, raw signaling data results and network structure are not shown which makes it difficult to make general conclusions of future possible performance. In order to do that the cause of degradation has to be isolated, e.g. location inaccuracy, wrong route determination or simply to few samples in a time interval. Simulation models can also be a good way of analysing expected performance of these systems, both in general and for specific geographic locations. Several simulation studies have been performed based on the active monitoring approach. Since recent systems are focused on passive monitoring, there is
a demand for passive monitoring simulation models that models the cellular network dynamics in greater detail.

Future systems will be fed with more information from data communications, rather than telephone calls. This part of the research area is surprisingly unexplored and development is needed in order to utilise the increase in data traffic. The shift towards UMTS will also affect the performance in these systems. In the coming years there will be many countries with a combination of UMTS and GSM terminals. Most likely the share of UMTS terminals will grow. In Sweden, for example, the number of new UMTS terminals has passed the number of new GSM terminals. Hence, using also UMTS terminals for travel time estimation will increase the number of floating sensors in the system, which is an important factor when it comes to both travel time accuracy and time to incident detection. Increasing the number of vehicle samples in a travel time reporting interval is very important for accuracy, especially since the system produces relatively noisy measurements.

The higher data rate and shorter delay together with dynamic measurement reporting in UMTS makes the network react much faster to changes in the radio environment, which affects the location accuracy of network events. This, in combination with the soft handover principle that makes a radio link addition or removal without large delay, might be the reason to the much better UMTS location accuracy in the evaluated tests. Generally speaking there is also a better synchronization between base stations and mobile terminals in UMTS and (eventually) a more dense radio network which gives a potentially higher location resolution. This is independent of whether handovers or something else is used to determine specific locations of the mobile terminal on the road.

The higher location accuracy in the UMTS network can be used to make the travel time accuracy better or maintaining the relative accuracy while making the travel time segments shorter. Shorter travel time segments are necessary e.g. in urban environments and are also useful when detecting incidents.

Relying on people using their cell phones in order to retrieve road traffic information is not always a perfect solution. It is likely that hybrid approach solutions will gain in popularity if a good way of selecting the cell phones to actively monitor is presented. The UMTS network utilising event-triggered measurement reports gives a new dimension to the hybrid monitoring approach. The location data available from a cellular network is highly dependent on network configurations. So far there has been little incentive for the operator to adjust the cellular network to perform better in road traffic estimation. This might change as soon as the demand for traffic information is high enough. An interesting analysis of this topic can be found in [60].

Most vehicles today have at least one cell phone onboard and a lot of valuable data is possible to extract from the cellular networks. If it is used in combination with other sensors or as a standalone technology will differ between different deployments. However, it would be a major waste not to use it at all.
REFERENCES


[17] Zhao, L. and Mark, J. W. Mobile Speed estimation Based on Average Fade Slope Duration. IEEE Transactions on Communications, Vol. 52, No. 12, pp. 2066-2069


