Road Traffic Tracking by using Location Management data in wireless cellular networks (TTLM)

K. Kyamakya, S. Fuchs, R.A. Koteswara, J.C. Chedjou
University of Klagenfurt, Group of Transportation Informatics, Klagenfurt, Austria
Email: kyamakya@isys.uni-klu.ac.at

Abstract

The need for increased individual mobility has resulted in a significant and continuous increase of road traffic density. More and more, the roads are getting congested. Modern Intelligent Transportation Systems (ITS) are offering methods to avoid these congestions. However, this requires a dense surveillance system of sensors that is very costly. Thus, due to financial constraints, not all roads are equipped with such sensors. Especially the secondary and tertiary roads are generally sacrificed. Thus, a complete real-time traffic monitoring of all available road systems is not guaranteed. This calls for new and flexible tracking systems which offer full coverage, high flexibility, synergy with available or already implemented systems and all this at a very high performance to cost ratio. Such a novel solution is presented in this paper. It uses location management data from wireless cellular networks and doesn’t require external sensors. The mobile phones, which have become ubiquitous devices carried by people moving inside the cars, are implicitly used as sensors. The key contribution of this paper is to show how cellular networks can perform a sensing task, in this case of road traffic information, while using functions they need for their normal operation, in this case the mobility management ones. Thus, the cellular network can be seen as a type of sensor network (even if not an ad-hoc one). This new service concept opens new revenue opportunities for cellular operators. The concept presented in this paper is named TTLM (an abbreviation for Traffic Tracking using Location Management functions of cellular networks).

1 Introduction

The number of vehicles on our streets is constantly growing. Today there are more than 53 million cars on German streets [1], responsible for an increasing number of traffic jams causing economical and ecological problems. The common German driver spends about 70 hours/year in traffic jams [2]. The lost time results in a waste of approximately 14 billion litres of gasoline. Not only Germany but every country has to face the same congestion problems concerning road traffic. The situation is even worse in highly populated countries like China although car ownership is lower there [8]. The fact that traffic density will continue to grow calls for efficient traffic guidance policy and systems to perform real-time load balancing that should result in an optimisation of the overall traffic flow through the available road infrastructure [9].

Figure 1 presents the key functions of a guidance system. Sensors and other sources of traffic information are measuring or monitoring current traffic conditions. The measured data are sent to a control unit which also has access to road infrastructure data and can map measured values to road segments. The subsequent processing produces a picture of the current traffic conditions. This information is then transferred to the drivers on the road. Basically, a traffic guidance system is set up to control traffic flows in order to reduce travel time. This is usually done by proposing alternative routes to avoid congestions or by using temporal speed limits to optimize traffic throughput. Because alternative routes are not mandatory for drivers, traffic flows can only be influenced within certain limits [12].

1.1 Road Network

The road infrastructure is subdivided in two parts: roads located outside and inside of cities. Roads of interest outside of cities are subdivided into two categories: federal highways and secondary/rural roads. Inside cities, roads are categorized into three subgroups: main axes, secondary axes and byroads. Outside, the road infrastructure is spread out, with long distances between the single roads. Traffic lights are just present in a limited number of locations [10].

Figure 1 Functional architecture of a road traffic guidance system
The main problem faced for traffic tracking outside of cities is the wide areas that need to be covered. A high density of sensors on these roads can only be achieved with high costs. Inside cities the road density is very high and there are a lot of traffic signals that interrupt a free flow [11]. The challenge here is the huge amount of different roads and bypasses. Consequently there are many alternatives to route through a city. Therefore, a highly dense network of sensors is needed in order to get enough data for reliable inference of the traffic conditions.

1.3 Traffic management systems

A global architecture for traffic management systems is shown in Figure 2. Traffic relevant data are collected by the sensors systems on the road or in cars and are sent to the traffic management centre. Additional information from police and other services can be involved, e.g. about accidents. The collected data are processed to generate a current picture of the traffic state and can be distributed to the customers and used for the control of road-side traffic guidance systems. This paper focuses on the generation and collection of traffic information and pays no attention to the distribution of information to car drivers or traffic guidance systems.

![Figure 2 General structure of a traffic management system](image)

The main challenge in traffic management is to ensure full coverage of all important roads, secondary roads and bypasses and not only focus on a limited number of hotspots. This is the crux in today’s traffic management. Secondary roads and bypasses are very rudimentarily covered by fixed sensors. Congestions on highways can be quickly detected due to the high density of surveillance systems. Car drivers taking the bypasses to avoid congestions often provoke a new traffic jam that is not detectable, because of the lack of sensors. The actual situation can be summarized as follows: a) the sensor systems don’t sufficiently cover secondary and bypass roads; b) full coverage of all road types is not realized due to high costs related to current sensor technologies; c) most of the current sensor technologies are too expensive or not able to provide accurate data in real-time.

1.4 Traffic Data

There are many different ways to track road traffic. The traffic data of interest in this paper can be classified as follows:

**Real-time traffic data:** These are collected by sensors and immediately transmitted to a traffic management centre. These data are essential in dynamic navigation as they are needed to recalculate alternative routes in case of traffic jams. Due to the high dynamic of traffic flows they can become out of date or obsolete very quickly.

**Street or road segments profiles:** When traffic data are recorded over long periods, profiles can be generated for the road segments, which consist of statistical values recorded per time unit, e.g. traffic density or velocity. By combining time and traffic conditions profiles can be set up to an hourly, daily, weekly or monthly profile. The temporal behaviour of road traffic does correlate e.g. the daily morning congestion on a main road artery through commuters. Profiles are also used for street planning and management. They are helpful for a first prediction of the traffic situation on a given road segment when real-time data is not available.

**Traffic forecast:** Forecast is based on microscopic, macroscopic and mesoscopic models. The microscopic model considers individual cars [21]. Every car is described by a set of rules describing the behaviour of the motorist. These rules can be e.g. the tolerable distance to the next car on the road or the average speed. From a macroscopic point of view [22] a traffic flow is analogue to a fluid flow and thus can be described by well known partial differential equations [22]. Mesoscopic models are a combination of both. With these models road conditions can be estimated by extrapolating from known history and/or real-time traffic flow data to a future traffic state.

2 Road Traffic Theory Basics

Different combinations of velocity, traffic density and occupancy result in different traffic states [14]. When only few cars are on a road, one speaks of a “free-flow”. Traffic density is low in this state. Motorists can decide how fast to travel within the admissible speed limit. With increasing density drivers have to pay attention to other cars. Vehicles start interacting with each other. They form platoons and the average distance between the cars is decreasing. As speed decreases more cars start hindering each other. Individual car speeds and distances between cars get unstable but traffic is still flowing. A further increase of density results in a more instable traffic flow. Traffic gets stuck due to extreme reactions like abrupt breaking. Stop and Go are alternating and can lead to serious congestion. Several parameters are used to identify traffic congestion: average speed, traffic density and traffic intensity. For further information about the interdependence of these three, see Refs [14, 15, 23].
3 Existing Sensor Systems

During the past several types of sensing devices have been developed. The sensors detect cars in a specific way, count them and sometimes perform a rough classification of vehicles. Some sensors measure at a point [16] others over a long section [17]. The widest spread sensors are inductive loops hidden in the road asphalt, others are radar, infrared, piezoelectric and ultrasonic sensors. Except for radar, velocity can not be measured “at a point” by most of the sensors. Two sensor measurements are needed at different points to calculate the average speed from the time interval between them. Systems using video images can measure traffic along the length of a whole segment [24]. Some tracking systems use satellite images or video cameras to evaluate traffic. With a single picture traffic density can be estimated. With several frames a single car can be tracked with image processing methods. Thus speed becomes measurable over the observed area but only with huge processing efforts.

Another approach is the Moving Observer Method (MOM) used in the Floating Car Data systems (FCD) [3]. MOM uses single cars as mobile sensor floating in the traffic. The observed car will statistically behave like an average car of the traffic stream. Thus the average speed can be inferred and even the car density if there are enough sensor cars. Cars equipped with FCD transmit their GPS-position to a server at regular intervals. At the server the travel time is estimated and the current traffic conditions are inferred. Data can be sent to the server via SMS, GRPS or over analogue wireless network (e.g. analogue radio systems used by taxi fleets).

Systems based on tracking mobile phones work in a similar way. Varying proposals with different accuracy levels have been suggested in the past years [25]. In principle they use the cellular network infrastructure for positioning the mobile phone. The simplest method is based on the received signal strength values. Other proposals use signal propagation time of arrival (TOA) or angle of arrival (AOA) methods. Today the penetration of mobile phones is big enough to provide sufficient probes. This section presents a new tracking system we introduced a set of criteria: coverage, flexibility, hardware requirements, synergy with other systems and cost. Table 1 summarizes the key features of the most important sensor systems. The quality based comparison shows the limitations. The ideal system should be optimal for each of the criteria. None of today’s traditional systems satisfies all formulated criteria.

<table>
<thead>
<tr>
<th>Sensor System Type</th>
<th>Coverage</th>
<th>Flexibility</th>
<th>Hardware</th>
<th>Synergy</th>
<th>Estimated Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radar Sensors</td>
<td>very bad</td>
<td>bad</td>
<td>bad</td>
<td>bad</td>
<td>bad</td>
</tr>
<tr>
<td>Inductive Loops</td>
<td>very bad</td>
<td>very bad</td>
<td>very bad</td>
<td>bad</td>
<td>very bad</td>
</tr>
<tr>
<td>Infrared Sensors</td>
<td>very bad</td>
<td>bad</td>
<td>bad</td>
<td>bad</td>
<td>bad</td>
</tr>
<tr>
<td>FCD</td>
<td>bad/fair</td>
<td>good</td>
<td>fair</td>
<td>good</td>
<td>fair</td>
</tr>
<tr>
<td>Tracking Mobile Phones</td>
<td>fair/good</td>
<td>good</td>
<td>very good</td>
<td>very good</td>
<td>bad</td>
</tr>
<tr>
<td>SAR</td>
<td>very good</td>
<td>excel-lent</td>
<td>excel-lent</td>
<td>excel-lent</td>
<td>excel-lent</td>
</tr>
<tr>
<td>Ideal Sensor System</td>
<td>excel-lent</td>
<td>excel-lent</td>
<td>excel-lent</td>
<td>excel-lent</td>
<td>excel-lent</td>
</tr>
</tbody>
</table>

Table 1 A qualitative performance comparison of the different sensor systems

The challenge is to design a sensor system which approaches the “ideal system”. Satellite radar imagery and FCD appear to be the best systems so far.

4 TTLM – the novel concept

This section presents a new tracking system we have developed with the intention to be as close as possible to the ideal sensor system. The new traffic tracking approach builds on synergy with existing systems and technologies, namely on the use of location management data that is already available and needed for other specific purposes in cellular networks.

The geographical coverage area of a cellular network is logically partitioned into location areas, which consist of groups of cells. Each cell is served by a base station. Mobile users are connected to the network via the base stations. Several base stations are connected to a base station controller and a number of base station controllers are connected to a mobile switching centre [30, 31]. Mobility in a cellular network is categorized in two areas: (a) Radio mobility which consists of the handover process and (b) Network mobility which consists of location management and is influenced by both user mobility and call patterns [30, 31]. Location management tracks the location of users’ phones inside a network. The granularity of this tracking has two levels [30, 31]: (a) between calls tracking is performed by the location update (LU) procedure at the location area (LA) granularity; (b) at call arrival the tracking is performed inside the location area by the paging procedure at the cell granularity. If the serving cell is changed during a call the cell selection is handled by the handover process. All data related to a call are recorded in appropriate data-
bases. To route an incoming call to a mobile station (MS) the position (LA) of the MS is recorded in large distributed databases, the so called Visitor Location Register (VLR). The data stored in the VLR is updated if a MS physically crosses into another LA [28]. The information provided by the VLR is one of the key pillars of the new approach in this paper. The Base Stations (BS) broadcast the Location Area Identifier (LAI) that uniquely identifies every LA. By crossing the border between two LAs a MS gets a different LAI and the LU procedure is initiated. Intra-VLR updates are done between two LAs located inside the same VLR. If the two LAs belong to different VLRs the corresponding LU is an inter-VLR update. Both the old and the new LAIs are known to the VLR. The following example explains the basic principle of the new sensor system. A LU is initiated at time \( t_0 \) as the MS enters the location area LA2 (coming from LA1); the transmitted data will be recorded in the VLR and the LAI will be changed.

**Figure 3** Illustration of the triggering of a location update

At time \( t_1 \), the vehicle passes the border between LA2 and LA3. The difference \( t_1 - t_0 \) is the travel time (see Figure 3) to a possible travel route by combining travel time, LU information and topological road information. We have to consider two types of networks: the cellular network and the road network. To obtain useful information maps of the cellular network and the road network are combined. If this is done, one can infer on the place (road segment) where the LU occurred. It’s important to distinguish between a LU performed by a mobile phone inside a car or a mobile phone anywhere else. This problem is handled in one of the next sections.

## 5 A probe-based Approach

Mobile phones moving within cars represent the quantity of traceable cars and have to be separated from the other ones. We focus on GSM phones, as GSM is the most popular system in Europe, but the assumptions and conclusions made here can be extrapolated to any other cellular phone system. In Germany there are about 65 Million GSM subscribers [29] and about 53 million cars. Fortunately, not all available data of the four cellular network providers is needed in order to get useful results. The roads of interest are the crowded ones and just a small portion of the cars present on roads are needed to calculate e.g., the average speed. In contrast to other sensor systems TTLM is a probe-based approach. It reduces the processing effort and offers the possibility to control the level of accuracy by considering more probes. The basic idea is that a probe will represent certain traffic characteristics sufficiently. A participating car will statistically behave like the average traffic stream. In this research field a lot of investigations have been conducted, e.g. [6]. They have shown that about 5% sample of all vehicles entering an area are enough to make a good estimation of the traffic state. The investigation in [6] also came to the conclusion that the information quality increases with traffic density, because there are more probes on the road. One challenging issue is the optimal selection of the 5% samples. Selecting arbitrary among the available mobile phones in vehicles would be inefficient.

**Figure 4** Schematic diagram illustrating the probes’ movements

### 5.1 User Mobility

**Figure 4** is the abstract scheme of a LA showing the mobile station flows. We observe the LA for a given time observation window depending on the needed update rate for traffic information. There is a number \( N_0 \) of mobile phones (probes) entering the LA; leaving probes are marked with \( N_1 \). In addition, there are some probes \( N_2 = N_0 - N_1 \) remaining in the LA. We should be able to distinguish between probes that will be crossing the LA and probes that will stay in that LA with high probability. The phones remaining in the LA (\( N_2 \)) have to be excluded, since the travel time can not be estimated if a phone does not cross the LA. If the exclusion is good enough \( N_0 \) should be very close (ideally, equal) to \( N_1 \). The mobility of most people can be seen as a movement between few activity locations only [31, 7]. These are work, residence, recreation etc. (in [7], they are called “Profile Nodes”). Most People spend the major part of the time in such locations. Often, there are temporal recurring patterns in the users’ movements, e.g. commuters driving to work every morning. This knowledge can be used in the strategy to lower \( N_2 \) to improve system performance. Consider the issue of separating MSs carried by pedestrians from those in cars. The key to solve this issue is to look at either recent movement history and/or at the observed average
speed. A mobile phone is only considered in the probe when it has been recently booked in a LA located far away in the recent past. This will help to exclude pedestrian mobile stations as pedestrians can not cover long distances in a short time.

6 Use of movement profiles to enhance performance

A user movement profile describes users’ movement history. These profiles are useful where exclusion or a differentiation of user types will be necessary. From a technical point of view creating a user profile is easy. Based on the recorded movement history (from the location management data) the network provider can generate a list of “Profile Nodes” for each user by tracing and extracting the LAs where a user regularly stays longer than a certain time threshold (network-based approach). Another way is a client-based application running on the mobile phones SIM card. This application could store the visited LAs and generate a list of Profile Nodes which can be transmitted to the provider at regular or occasional times. One should keep in mind that user profiles always involve privacy issues. In our traffic sensing application, however, user identity is not important. An appropriate algorithm can make the data anonymous while keeping them still unique. So doing privacy concerns can be calmed.

To select the probes for a particular area the provider doesn’t need to look up the complete database. He can first exclude all users having a profile node in that area. For such users the probability is high, that they will stay in the area and thus not contribute to the sensing task. This strategy will ensure that the component N2 (as indicated in Figure 4) is reduced.

6.1 Data Processing

After probe selection the corresponding location update data will be transmitted to the TTLM server, where further processing is performed. The transmission between the VLRs and the TTLM server occurs over the wired backbone network. An ID containing a user’s encrypted IMSI (International Mobile Subscriber Identity) is assigned to each element of the selected probe, thus ensuring an anonymous but unique tracing. The travel time differences are computed from the received data and, by using the mapping information, an assignment to specific road segments can be done. Knowing the segment lengths and the travel times, the average speed can be computed. A look at the average speeds and the speed variance helps determining the traffic state on a particular road segment. The best case is when the average speed is high and constant (“free flow”). If the speed drops below a given threshold the road segment enters the next traffic state (“partially bound traffic”) and so on. After a certain threshold the probability of having a traffic jam increases and finer sensing is needed. In our concept, we have another granularity level of probes, the cell level. A finer granularity also needs more signalling resources for the corresponding location updates.

6.2 Difficulties and ambiguous situations for the mapping

If two roads originate in the same LA and have a common border to another LA they can’t be distinguished from each other by just using LU information at the granularity level of LAs. A possible solution is an application in the phone (for e.g. in SIM card) which stores the IDs of all cells crossed inside the LA and transmits them during the next LU when entering a new LA. What happens if a highway is parallel to a local road in a LA? Another similar situation is a railway parallel to a road or a highway. How to distinguish probes in the train from those on the road? For such situations the short time history of affected probes has to be considered to exclude ambiguity. In both problematic cases probes can be separated by looking at both average and variance of their speeds. Besides, the sequences of the previously visited LAs should be investigated. For a set of probes moving in a train the sequence will be the same.

6.3 Escalation Levels

It is possible to scale the accuracy by changing the amount of probes involved in the sensing process. Escalation steps can be defined e.g. by using a modified and more accurate location update policy (LUP). The escalation levels make the system flexible and reduce the necessary signalling load. Fast reactions to changes are possible without generating additional global load. Higher escalation levels should contain the lower ones. Implementing a specific application on the mobile phone (in the SIM card) allows the setting of the levels and datatypes to be transmitted during the LU process. The first escalation level uses the information generated by the LUP of the GSM standard (“normal” LUP). The information contains the previous and the actual LAs and the related cross-over time. In most cases this is sufficient to determine the traffic state. The second LUP level requires transmission of additional information. The cells visited inside a LA are stored on the SIM card. The sequence of visited cells will indicate a probable route within the LA with more confidence. The third level also determines the time spent within each cell which is recorded by the mentioned SIM application. This information will be transmitted along with that of the first and second levels in the next LU event. Thus, congestion localisation in problematic cells is more accurate. The third level is also useful in distinguishing mobile phones in cars from those in trains. The fourth level requires intensive monitoring of one or more problematic cells (with high probability of congestion). For this purpose, a new LUP strategy is needed, which forces a mobile station to perform a LU of visited cells and LAs in given time intervals, regardless of changes of its
location. The LUP is initiated by the application in the mobile phone as the selected mobile phone enters the problematic area. LUs will be generated at regular time intervals until the mobile phone leaves the problematic area. At this level the LU information will also contain the actual cell ID and compressed additional information like the signal strength of the serving and neighbouring cells. Based on this information the TTLM server can make an estimate of the mobile phones location which allows better tracking of the phones movements. In the fifth escalation level single probes can be actively located with an accuracy between 50m and 300m by using a precise positioning system implemented in the cellular network (E-OTD, AGPS, etc. [25]). This is however expensive as it requires considerable load on the signalling channels. During a call localisation is less complicated, since timing advance values are known to the network. This value is a rough measure for the distance between a mobile phone and a BTS. In the fifth phase phones in conversation mode should be selected as probes with priority. This escalation level should be used only in special situations to resolve a complex ambiguity when all lower levels were not successful. When the phone leaves problematic LAs the escalation level can be automatically reduced. Level 1 should be the default as it is normally adequate for most situations.

7 Architecture of the TTLM System

The architecture of the TTLM system is presented in Figure 5. The system can request data from network’s registers. Requested data are from the HLR in case of intra-MSC handovers and from the VLRs in case of Inter-MSC handovers. TTLM can also initiate escalation level changes. Beside requested data and orders to change escalation levels no information is transmitted from the TTLM to the network, thus minimizing needed data rates. A single network provides the requested data in real time. Other networks can be easily connected to improve performance by providing a larger set of probes. The data traffic quantity between the network and the TTLM server depends on the probe sizes and the escalation level. TTLM only uses a small portion (some percent only) of the data available in the location management databases. This saves communication resources both on the air interface and in the backbone wired network [4].

8 Conclusion

The quality of navigation decisions based on announced traffic information highly depends on the accuracy and on the actuality of the sensed tracking data. Up-to-date traffic conditions are essential to guarantee the functionality of a global traffic management system. This paper has first presented an overview of the state-of-the-art in real-time road traffic information sensing. With help of some evaluation criteria it could be shown that current systems are far from ideal.

Figure 5 Server and communication architecture of the TTLM system

A new traffic sensing concept developed with the intention to be as close as possible to an ideal one has been presented. It uses the cellular networks, which generally have high penetration and area coverage to generate up-to-date traffic information at a very low cost. A comparison with other systems has shown that they are either more expensive or less flexible. The basic idea of the system is to use data that is already available for location management in cellular networks. Different escalation levels are used for fine location management resulting in a lower signalling load for the paging process for the probes concerned. User profiles and recent movement history can be used to solve a series of possible ambiguity cases. Another significant advantage of this solution is that it is purely a “software solution”\(^1\). No additional hardware is needed on the roads. Furthermore this new service does not disturb the normal operation of the cellular networks and it is a great source of additional revenues for the network providers. An ongoing work is developing a coupling concept between the “cellular network mobility and billing functions” and a “floating-car data sensing infrastructure” (involving GPS for positioning if available or a car-to-car communication ad-hoc network). Such a system will probably ensure a very robust and cost-effective traffic sensing system.

9 References


\(^1\) Note: a patent application related to TTLM has been submitted to the German patent office.


[18] ETSI Technical Standard 100 530, GSM 03.12 Digital Cellular Telecommunications System (Phase 2+), Location Registration Procedures.

