Road Traffic Monitoring Systems

Seminar: Datenkommunikation und Verteilte Systeme
SS 2003

Andriy Panchenko
Matrikelnummer: 236400

Betreuung: Dipl.-Inform. Karl-Heinz Krempels
Lehrstuhl für Informatik IV, RWTH Aachen
## Contents

1. **Introduction**  
   - 3

2. **Positioning Technologies**  
   - 2.1 CGI+TA (Cell Global Identity + Timing Advance)  
   - 2.2 UL-TOA (Uplink Time-To-Arrival)  
   - 2.3 E-OTD (Enhanced Observed Time Difference)  
   - 2.4 A-GPS (Assisted Global Positioning System)  
   - 4

3. **Geographic Data Files**  
   - 5

4. **Global Positioning System**  
   - 4.1 Introduction  
   - 4.2 GPS basics  
   - 4.3 DGPS basics  
   - 7

5. **Traffic flow monitoring**  
   - 5.1 Detailed description  
   - 5.2 Use of Traffic Monitoring System  
   - 10

6. **The Traffic Estimation Model**  
   - 17

7. **Summary**  
   - 19

References  
   - 20
1 Introduction

As the road congestion has increased over the past 20 years, there has been a corresponding increase in the need for a high quality, real-time road traffic information. Many approaches have been taken to gather this information, but fundamentally one can divide them into the following three groups:

- **Simple observation.** A picture about the road conditions is build up by gaining the information from such sources as callers to the radio stations, observations from the sky helicopters and emergency services reports. This approach suffers from the complexity of the integration of these sources and the accuracy they provide. Timeliness is the subject of doubt and the coverage of the information can be suspect.

- **Fixed sensor deployment.** Observation is made by some roadside sensors, which typically either measure the speed of the traffic on some predefined points of the road or track the individual cars between carefully selected way points. The observed data is then sent to the central site for real-time analysis. The granularity and timeliness of the information is very much dependent on the intersensor spacing. However, this approach is expensive both to deploy and maintain. Example of such a system on the German Autobahn network has been deployed by Tegaron.

- **Floating car data (FCD).** To overcome the drawbacks of the two above mentioned approaches, there has been a great increase of the interest in using the travelling vehicles themselves as sensors. But in order to do this on the cars some additional equipment has to be installed. This is typically GPS receiver and a sender, which would transmit current position to the central site. Unfortunately the fleet management market is fragmented, resulting in small, statistically insignificant pools of probe vehicles giving poor road network coverage. But a number of companies have correctly observed, that most of the vehicles already carry a sensor on board - the mobile telephone belonging to the driver or passenger. It should be mentioned, that the state of the art in locating subscribers lays behind the general mobile telephone networks consideration. Moreover, any approach which involves locating a subscriber places a significant load on the network. Therefore it makes more sense to monitor the network passively for location data with the help of normal subscribers activity.

However, a single cell can cover an area up to $4000km^2$ and such estimation is clearly not applicable for the purpose of road traffic monitoring. In the networks based on the GSM technology a further refinement is possible. Whenever a handset is in call, the network has to compensate for its distance from the cell site to ensure that signals from all handsets arrive at the base station at the correct time. This compensation is known as the timing advance (TA), and can be used to narrow a handset’s location to a approximately $550m$ wide concentric band radiating from the cell. However, also using this feature in the worst case there is still an area of uncertainty of $100km^2$, which is also clearly not enough for finding out which road a driver is travelling along.

By analyzing a sequence of TA reports from a particular handset over a short period of time, its route and speed can be determined. At the time of the first report the system knows that there is a
subscriber in some region. As further reports arrive, the system identifies the handset as moving and performs a detailed analysis of the possible routes that may have led to such a sequence of reports. This approach compares favourably with the deploying a fixed sensor network with sensors placed every 550m one from another. Indeed, in areas with a dense mobile network coverage the overlaps between the adjacent cells serves to improve the accuracy to the order of a few hundred meters.

By gathering route and speed data for many active subscribers, a statistically significant view of the road situation can be established. Moreover, this data can be gathered without making expensive additions to or using the existing capacity of the mobile network. The behaviour of subscribers works in systems favour. When a road is lightly used and free flowing one needs few reports to declare such a situation, but when congested, there are typically more drivers on the road and hence the data available to the system increases.

The article begins with a brief description of the available positioning technologies. It than reviews the Geographic Data Files to represent road networks. Global Positioning System basics are covered in the chapter 4 and the Traffic Flow Monitoring System description in section 5. As another possible traffic estimation technology the Traffic Estimation Model will be presented. Finally, in the Summary, the review discusses comparison of both presented approaches.

2 Positioning Technologies

The principal positioning technologies available for the mobile subscriber device location include the following:

2.1 CGI+TA (Cell Global Identity + Timing Advance)

This method can determine the distance of an active Mobile Subscriber (MS) device (i.e. one currently engaged in a transmission) from a particular base station to an accuracy typically of the order of 550m (within an annual zone (complete 360°arc) around the base station which has a radial depth of 550m). The information can also be gained by “paging” an “idle” handset (i.e. one which is switched on but not currently engaged in a call). This method requires no MS device modifications. A base station with multiple directional antenna (which are now common) reduces the location arc to a sector around the base station of, for example, 120°. Further enhancements are planned to increase the accuracy of this method to between 100m and 200m.

It should be noted here, that with some kinds of networks, for example GPRS (General Packet Radio System) networks, an MS device which is switched on but not actually involved in sending anything to or from the MS, is still in communication (at least periodically) with the call management system for the purpose of managing the network and, therefore can be considered as an “active” MS device in the context of the broadest scope.
2.2 UL-TOA (Uplink Time-To-Arrival)

UL-TOA can determine the location of a MS to the area within the 50m to 150m depending upon the terrain, by measuring time taken by the signal from the mobile handset to arrive to multiple “measurement points”. In other words - distances from each of these different measurement points, determined from the respective times, can be used to determine the position of the MS device.

2.3 E-OTD (Enhanced Observed Time Difference)

Unlike CGI+TA and UL-TOA this method places the responsibility for determining location in the MS device, and hence incurs little extra expenses for the mobile operator. Essentially this approach is the reverse implementation of UL-TOA. The accuracy is similar to that of UL-TOA (about 60m in rural areas and 200m in bad urban areas).

2.4 A-GPS (Assisted Global Positioning System)

GPS is commonly used for navigation systems in cars. This technology relies on the network of satellites orbiting the Earth and transmitting signals which can be received by a unit on the ground and used to calculate its own location. The GSM network can provide assistance that gives increased accuracy over standalone GPS systems by making use of the actual precisely known position of the base stations as reported by the GPS system in order to generate a correction factor which can be applied to the mobile subscriber device position as reported by the GPS system. The accuracy of this method is extremely high but requires modifications to mobile handsets.

3 Geographic Data Files

The road network data used for the traffic monitoring systems is generally in the form of a data file which can be more or less easily operated on mathematically. One convenient readily available and adaptable data file format is GDF (Geographic Data Files) [6]. It is a European standard, that is used to describe and transfer road networks and road related data, in which road networks are stored in the form of nodes representing junctions and edges representing each carriageway or road direction between neighbouring junctions. It is much more than a Geographic Information System (GIS) generic standard, because GDF gives rules how to capture the data and how the features, attributes and relations have been defined. Moreover, this particular data file format has the advantage that it can include information on the classification of roads i.e. distinguishing between motorways and other major or trunk roads and minor roads, which can be used as a basis for weighting such roads when constructing a probability vector for a vehicle on the basis that there will generally be a greater likelihood that a vehicle is traveling along a major road than a minor one. GDF has been developed in
a European project namely EDRM (European Digital Road Map). It’s primary use was for car navigation systems (Bosch, Philips, Volvo etc.), but it is very usable for many other transport and traffic applications like Fleet Management, Dispatch Management, Traffic Analysis, Traffic Management, Automatic Vehicle Locations etc. At this moment, GDF version 3.0 has been released and issued to CEN (Central European Normalisation) for the voting procedure. After the voting GDF will become the only CEN accepted standard for digital road networks.

GDF has a three level structure. These levels should not be considered as completely separated representations, but as different structures in one and the same representation, the higher ones embedded in the lower ones. However, these structures are all transparent in the sense that they all have a meaning of their own.

- **Level 0 Topology.** This is a common GIS topology description that is widely accepted. i.e. everything has been described by nodes, edges and faces. Nodes and intermediate points are represented by exactly one pair (XY) or triplet (XYZ) of coordinate values. A segment is bounded by exactly two intermediate points and/or nodes. A chain contains always one or more segments and is always bounded by exactly two nodes. Nodes and chains together form a planar graph.

- **Level 1 Features.** Level 1 is the most used level of GDF. It contains simple features like road elements, rivers, boundaries, signposts etc. Features can have attributes that are specific to the feature. i.e. one way, width of the road, number of lanes. Features can also have relations. These relations are very important for car navigation systems. Relations can be ”forbidden turn from road element 1 to road element 2” or ”road element 1 has priority over road element 2”.

- **Level 2 Complex Features.** At this level the ”simple features” are aggregated to a higher level feature. For instance at level 1 all road element of an intersection should be represented. At level 2, the intersection is only represented with a single point.
4 Global Positioning System

4.1 Introduction

The problem of navigation has appeared as humans began to travel. On the government level supported navigation systems are already for years used for airplains and ships. They were primarily based on the “radar” technology, measuring the travel time from the on earth positioned station to the traveling vehicle and back. The very first ideas of current GPS system have appeared in the late seventies, but the official trial was lunched only in 1993. Altogether it took about 6 years to develop the system and 8 to build it. It belongs to the oddities of the militar history, that americans and russians have developed believed to be independent from each other at the same time very similar navigations systems. They are so similar that in further developed twin-devices for american GPS and russian Glonass [11] some units were usable for both systems.

4.2 GPS basics

21 satellites (plus 3 in reserve) circuit on the height of 20 200 km above the globe. This is not a geostationary position, but rather the satellites browse the sky like extrem high flying airplains. Their paths are calculated extremely precise and distributed such a way, that at any place on the Earth at any time it is possible to “see” at least four of them. On the board they have two transmitters with frequencies $L1 = 1575,42 MHz$ and $L2 = 1227,60 MHz$, two cesium-atom-clocks and two rubidium-atom-clocks. Only atom clocks provide time to an accuracy of nanoseconds. Different frequencies are needed in order to eliminate errors coming from the ionospheres refraction of the waves and as the second reason - to provide two different precision-classes for localization. Different clocks reduce measurements error and improve short- and long time constants. The transmitter modulates two different long impulses sequences. One of them has the repeat rate of about 1 MHz, another of 10 MHz. So the last one is much more shorter, or in other words: in short time there will be much more impulse sequences transmittet. The first PRN-Code (Pseudo Random Noise) is a common knowledge and is used by a traditional GPS-receiver to locate itself with the accuracy of about 30 meters (so called C/A Code = Coarse/Aquisition-Code).

The second code is one of the top secrets of US Army and is known only to a limited number of people. Using this P-Code (Precision-Code) it’s possible to make localization with the accuracy of 3-5 meters. In PRN-Code there is also identification of the satellite presented, so codes of different sattelites are unique. After being switched on, the GPS receiver demodulates signals on the wideband frequency L1. These signals are nothing else as sequences of zeros and ones. But receiver also knows some predetermined sequences, actually C/A-Codes, with the help of which like a pattern, repetition rate of 1 MHz can be determined. The receiver tracks it narrowbandly further. Together with the individual PRN-Codes each of the satellites transmits further data, but pretty slow, with a speed of only a few bauds. In detail these are so called ephemeris (tracking-data, corrections for the ionospheres irregularity, status information of the satellites, time and path correction). In order to read all those informations from each satellite some amount of time is required. The last depends on
the sensivity of the receiver and some other parameters and can take up to 20 minutes. This data will be saved (also when the receiver is being switched off) for the future use and refreshed regularly as the new information becomes available. In the next the receiver knows the exact time and position of the satellite. Parallel to the receiving of the ephemeris further satellites PRNs being received and tracked. After this initialization phase localization can begin. For this reason the signal travel time from each of the satellites to the receiver has to be calculated. Knowing the speed of light, the distance can be determined. So having one measure like above mentioned, lies the position on the sphere with the center at the satellite and the radius of the distance. With two measures the possible position is reduced to the circle-like curve. The third one (third sphere) leaves only one possible (close to Earth) common point for all of the spheres. Nevertheless, there is also one blemish: clocks in satellites and in the receiver are going not exactly synchron. Time difference of only 1 μs causes positioning failure of about 300m. With the help of the fourth satellite this inaccurac y will be eliminated. Mathematically this idea can be represented as the system of four equations with four unknowns:

\[
\begin{align*}
    \begin{cases}
        p_1 &= \sqrt{(x_p - x_1)^2 + (y_p - y_1)^2 + (z_p - z_1)^2} + \Delta p + e_i \\
        p_2 &= \sqrt{(x_p - x_2)^2 + (y_p - y_2)^2 + (z_p - z_2)^2} + \Delta p + e_i \\
        p_3 &= \sqrt{(x_p - x_3)^2 + (y_p - y_3)^2 + (z_p - z_3)^2} + \Delta p + e_i \\
        p_4 &= \sqrt{(x_p - x_4)^2 + (y_p - y_4)^2 + (z_p - z_4)^2} + \Delta p + e_i
    \end{cases}
\]

where

\[
\begin{align*}
    &p_1 \text{ to } p_4 \quad \text{pseudo ranges} \\
    &x_p, y_p, z_p \quad \text{actual coordinates of the user} \\
    &x_i, y_i, z_i \quad \text{coordinates of the corresponding satellite i, } i=1,2,3,4 \\
    &\Delta p \quad \text{inaccuracy of the distance from the satellite} \\
    &e_i \quad \text{additional inaccuracy in the system}
\end{align*}
\]

Figure 2: Intersection of three spheres
After the first localization was done (so called “fix”), ephemeris will be saved and updated as soon as new information becomes available. If the receiver was shortly switched off or unable to receive the signal (e.g. in tunnel) GPS device will know PRNs of which satellites should it be possible to receive in a concrete moment. Until the end of the cold war the accuracy of the civil GPS devices was artificially limited (so called “selective availability”). Very slow modulation of the clocks took care for a mistake, which was typically about 200 meters in south-north direction and about 100 meters in west-east. On May 1, 2000 “selected availability” was switched off and since that time inaccuracy lies in 95% of the cases within a circle with the radius of 12 meters.

The output of the GPS device is usually in so called NMEA-Format (National Marine Electronic Association) with a serial RS232 interface. Nice is that the information is transmitted in plain-text and it’s possible to see it with any terminal programm.

4.3 DGPS basics

The idea of the DGPS (Differential-GPS), described in [10], is to use an additional base station on the Earth, which would calculate the inaccuracies sended by the satellites and would send correction data to vehicles e.g. thru LW (Fig.3). The Real-Time Differential GPS over the Longwaves was developed together with the Institute for Applied Geodesy Potsdam and Deutsche Telekom AG in 1995. The sender was installed in Mainflingen (near Frankfurt am Main) and sendet on 123,7 kHz in RDS (Radio Data System) encoding. With the help of the appropriate LW receiver this information can be transferred into the RTCM (Radio Technical Commission for Maritime Services) -capable GPS device through the RS232 interface. Real Time Differential GPS was able to improve the accuracy to
be not worse than 5 meters in radius and the sender was able to cover the territory a bit greater than Germany itself. But as the “selected availability” was switched off the need of the DGPS became not any more so important as before.

5 Traffic flow monitoring

The idea of the traffic monitoring system [2] is to capture geographical positioning data of individual active mobile subscribers and converting these into probability vectors representing the likelihood of the vehicle having arrived at any of the possible road components compatible with the geographical positional data. As the vehicle travels along, this process is repeated and new probability vectors are constructed. They are based on the probabilities of the corresponding routes taken between the current and previous location data. The expected transit time $\Delta t_x$ for the available routes are computed and compared with actual transit times $\Delta t$ to provide delay factors for the available routes and thereby the corresponding road components. Average delay factors are obtained by making use of the data gained from the other vehicles. With the help of the average delay factors the degree of the traffic congestion on the specific road component is evaluated.

5.1 Detailed description

In detail method comprising the following steps:
- capturing first geographical position for an active mobile device used on a vehicle at a given time \( t_1 \);

- intersecting first geographical position data with the road network map and defining original possible road components of the first geographical positional data;

- generating an initial probability vector representing the likelihood of the vehicle having arrived at a position on a given one of original possible road components for all possible road components;

- capturing second geographical positional data for the mobile device at a later time \( t_2 = t_1 + \Delta t \) where \( \Delta t \) is the actual transit time of the device between first and second geographical positions;

- intersecting second geographical position data with the road network mapping data, so as to identify new possible road components corresponding to the second geographical positional data;

- identifying available routes in the road network linking possible road components corresponding to the first and the second geographical positional data;

- generating an updated probability vector representing the likelihood of the vehicle having arrived at a position on a given new possible road components corresponding to the second captured positional data at the later time \( t_2 \) via one of the available routes for all of the new possible road components;

- intersecting available routes with expected average vehicle speed data for each of the series of road components in order to determine expected transit times for available route.

- directly or indirectly comparing the actual transit time with the expected transit times for each of available routes in order to produce delay factors for routes;

- determining an average delay factor for a plurality of vehicles, which is then weighted on the basis of the likelihood of any of the available routes have been followed.

The particular positioning or location technology used is in many respects unimportant for the method. The common attribute all these methods share is that the location position for each mobile device may be expressed as being within a given area of uncertainty. It is the responsibility of the system to “fit” a series of such readings onto a physical road traffic network and identify those readings which are likely to be in moving vehicles. The mobile telecommunications equipment vendors are developing various positioning solutions based on one or more of the above mentioned technologies. Most of these companies offer their own interfaces but there is an ongoing effort to standardise them. The Ericsson Mobile Positioning Protocol (MPP) has been selected as the basis for the standardisation. This provides an interface with which to query the Ericsson (or other compatible) Mobile Positioning Centre (MPC) in order to extract positioning data for individual MS devices. The MPP hides the particular mechanism which is used by the MPC to locate the MS device which therefore could be based on any of the before described technologies.
The size and form of the area of uncertainty defined by the positioning system may vary according to the particular positioning system used. In the case of CGI-TA each base station serve a sector-shaped area radiating out of the station, where the sector arc can be 360° or any smaller angle, such as for example 120°. The sector may extend several kilometers or more in any given direction depending on the topography of the area. Due to the increased delay observed in signal transmission between a station and a mobile device as the distance between them increases, the sectors are divided up into a series of annual timing advance zones. A vehicle, traveling away from the base station passes from one timing advanced zone to a neighbouring one in which signals are subjected to a different timing correction so that these delays can be compensated and the signals from various mobile devices at different distances from the cell are all properly synchronised. Typically the radial extend of each zone is several hundred meters, about 500, but may be up to 1500 meters or more depending on the network. There is generally a small overlap (intersection region) which may have a radial extend of the order of 50 to 100 meters. In order to facilitate the calculation of the expected transit time we make an assumption, that the vehicle is at the earliest part of the road component as a new timing advance zone report arrives. For the localization purpose it would be enough if the cell would transfer to the main site its identity and the particular timing advance zone, within which the mobile subscriber is located. This information can be then converted into coordinates. Road length units should not be too small, preferably about 500m. Thus, for example, if a road segment in the geographical data file was greater than 500m, then the data would be modified by breaking that segment up into shorter units, each of not more than 500m length. In the case of urban and sub-urban areas with a dense road networks, there will be very large numbers of very short road segments. In order to simplify and reduce the processing load it may be desired to treat several road segments as part of a single unit in such a case. In the case of freeways and other major highways with two (or more) separate carriageways, each of them is normally represented as separate road segment.

The generation of the probability vector representing the likelihood of the vehicle having arrived at a position, on one of the possible road components, may be done using any suitable criteria. Generally this is the classification of the road and, desirably, also the length of the road within the timing advance zone. Using the CGI + TA positioning system, the geographical positional data is generally captured when the device crosses a timing advance zone boundary between one timing advance zone and a neighbouring timing advance zone. Thus, the system initially generates a probability vector when a vehicle, carrying an active mobile device, crosses a first timing advance zone to a second timing advance zone. When a vehicle crosses a second timing advance boundary from the second timing advance zone into the third one, the system constructs a transition matrix representing all possible routes that could have been taken to get from the first timing advance boundary to the second one. For each route a probability is calculated. It can be determined by the product of the selected criteria, e.g. length of the road multiply by classification weighting. In addition an expected transit time is calculated based on the length of the road taking into account standard speed on the concerned road. The actual transit time between the crossings of the first and second timing advance boundaries may then be compared with the calculated expected transit time to provide an additional probability factor. The factor is based on the fact that it is significantly less likely that the actual transit time will be substantially less rather than substantially more than the calculated expected transit time. This additional probability factor may then be applied to the transition matrix to produce a time
dependent transition matrix, which can then be used to update probability vector. Thus, for example, where one or more of the originally available routes are absent from the time dependent matrix, then this can now be excluded from the updated probability vector. The collected information can be then combined with that for the other vehicles found to be using the same road component in order to provide an average speed. The average is customized to provide increased weighting for faster moving vehicles as these will be more representative of the maximum available rate of progress on that road and hence the degree of congestion at that time. The degree of congestion is determined by comparing the calculated average with a normal (uncongested condition) average speed. It can be then provided on any convenient scale, such as a numerical or percentage scale. By counting all the vehicles found to be using a particular road component, it is also possible to estimate the volume of traffic on the road (based on a typical proportion of vehicles carrying an active MS). This information can be then optionally used with other additional information such as time of day or night, weather conditions etc. in order to further refine the calculations. Thus, for example, the composition of the traffic in the middle of the night is likely to have a higher proportion of trucks (which are subject to lower speed limits than other vehicles) than during the day, which would result in the calculated average speed being wrongly decreased. Therefore the expected average speed used for comparison purposes at such times could be adjusted. Alternatively, the expected speed could be kept unchanged, and the weightings used in the calculation of the average speed at such times could be modified. It should be noted that up-to-date average speed and delay reports will depend on the frequency with which geographical positioning data can be captured. Thus, for example, in CGI+TA geographical positional data is captured when a vehicle crosses timing advance zones boundaries. The greater the separation between these and the slower the vehicle speed, the longer the interval between the capture of the geographical positioning data is. In practice such intervals can typically range from less than one minute to several minutes or more. With other positioning systems, such as, for example, A-GPS, positioning data can be captured rather more frequently and regularly, for example, at a fixed interval in the range from 5 to 30 seconds.

Real-time system subjected to a greater or lesser degrees of “noise” and there should be included suitable filtering to reduce the effects of such noise. Thus, for example, it can happen that a mobile phone instantaneously switches from one base station to another and then back again without having moved at all, due to fluctuations in relative signal strength of neighbouring base stations e.g. due to particular weather conditions etc.

The reports about road congestion may then be made available in a generally known manner through any suitable interface, including synthetic voice reports, graphical representations, text reports via SMS (Short Message Service), HTML (HyperText Markup Language) and WML (Wireless Markup Language), cell broadcast message format for transmission via CB (Cellular Broadcast) Centers, etc. The delay factors reports commonly would include at least some earlier reports, which have been suitably aged or decayed at a linear rate of 10% per minute for a busy road and 5% per minute for a quiet one.
5.2 Use of Traffic Monitoring System

When a vehicle (Fig.5) originally within the position defined by timing advance zone 100 is observed to enter the geographical position defined by timing advance zone 200 at time $t_1$, the initial probability vector $V_1$ is constructed for all the possible road components that the vehicle could be on - in this case those that lie in the timing advance zone labeled 150. It is important to mention that the system only begins capturing geographical positional data reports for a mobile subscriber device when it detects that it has changed its position from previous one and, thus, is moving, thereby filtering out reports for stationary subscribers who are unlikely to be in a vehicle driving along a road. The initial probability vector $V_1$ would have the following form:

<table>
<thead>
<tr>
<th>Route</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{2150}$</td>
<td>0.3</td>
</tr>
<tr>
<td>$A_{1150}$</td>
<td>0.4</td>
</tr>
<tr>
<td>$A_{6150}$</td>
<td>0.3</td>
</tr>
</tbody>
</table>

The probabilities are skewed by the type of the road: major roads or highways have higher probabilities.

At time $t_2$, the vehicle is observed to cross from timing advance zone 200 to timing advance zone 300. In order to construct the transition matrix $A$ for determining the updated probability vector $V_2$ representing the new position of the vehicle, the set of all possible route from timing advance zone 100 to timing advance zone 300 via timing advance zone 200 is determined by use of a route finding...
algorithm. Each route consists of a starting point on the inner edge of timing advance zone 200 (i.e. the edge closest to the base station), a set of road components within timing advance zone 200, and an ending point on the inner edge of timing advance zone 300.

The expected time take to get from starting to ending point and the probability for each route are also calculated:

<table>
<thead>
<tr>
<th>Route</th>
<th>Expected time (sec)</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2150 (\rightarrow) A7250</td>
<td>55</td>
<td>1</td>
</tr>
<tr>
<td>A1150 (\rightarrow) A1250</td>
<td>23</td>
<td>0.8</td>
</tr>
<tr>
<td>A1150 (\rightarrow) A5250</td>
<td>32</td>
<td>0.2</td>
</tr>
<tr>
<td>A6150 (\rightarrow) A6250</td>
<td>35</td>
<td>1</td>
</tr>
</tbody>
</table>

The expected time for the route taking only road A1 is much less than the routes using the more minor roads. The probability column represents the probability of using a particular route, given a particular starting point. Hence the \(A2_{150} \rightarrow A7_{250}\) and the \(A6_{150} \rightarrow A6_{250}\) routes are both given initial probabilities of 1, as there is only one possible route that could be taken given those starting points. But the two routes involving road A1 have different probabilities based on the type of the road (higher for road A1, lower for road A5), and given that the total (initial) probabilities of all the routes starting on A1 will be 1, then the probabilities for each of these (\(A1_{150} \rightarrow A1_{250}\) and \(A1_{150} \rightarrow A5_{150}\)) will be less than 1.

The transition matrix \(A\) can be cached for future use for vehicles in a similar position thereby reducing the computational processing load, as the transitional matrix is expensive to calculate. To use it in this case, it is first converted to a time dependent transition matrix. For each route, the actual transit time \(\Delta t\) is compared to the expected transit time \(\Delta t_x\) to provide a delay factor which is also used to adjust the probability of the route. Those routes, with expected transit time \(\Delta t_x\) significantly longer than the actual transit time \(\Delta t\) have their probabilities reduced to reflect the fact that it is unlikely that drivers in general will travel at substantially above the speed limit. (A suitable formula would be a linear reduction of the probability to zero for increased speed above the expected speed up to a speed which is double the expected speed.) Hence for an actual transit time of 30 seconds, the time dependent transition matrix \(A_{30}\) would look like this:

<table>
<thead>
<tr>
<th>Route</th>
<th>Delay Factor ((\Delta t/\Delta t_x))</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2150 (\rightarrow) A7250</td>
<td>0.55</td>
<td>0.65</td>
</tr>
<tr>
<td>A1150 (\rightarrow) A1250</td>
<td>1.3</td>
<td>0.8</td>
</tr>
<tr>
<td>A1150 (\rightarrow) A5250</td>
<td>0.94</td>
<td>0.15</td>
</tr>
<tr>
<td>A6150 (\rightarrow) A6250</td>
<td>0.85</td>
<td>0.9</td>
</tr>
</tbody>
</table>

The updated probability vector \(V_2\) is given by the product of the initial probability vector \(V_1\) and the route probability from the time dependent transition matrix \(A_{30}\). This is then normalized such that the sum of the probabilities is 1.0. (In more detail \(V_2\) is generated from the time dependent transition
matrix by multiplying the probability for each route by the probability of being at the starting point for that route as obtained from the previous (immediately preceding) probability vector $V_1$, this process being repeated iteratively.)

<table>
<thead>
<tr>
<th>Route</th>
<th>Probability before norm.</th>
<th>Probability after norm.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_2_{150} \rightarrow A_7_{250}$</td>
<td>0.2</td>
<td>0.24</td>
</tr>
<tr>
<td>$A_1_{150} \rightarrow A_2_{250}$</td>
<td>0.32</td>
<td>0.38</td>
</tr>
<tr>
<td>$A_1_{150} \rightarrow A_5_{250}$</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td>$A_6_{150} \rightarrow A_6_{250}$</td>
<td>0.27</td>
<td>0.32</td>
</tr>
</tbody>
</table>

It is now assumed that the delay factor for a route applies equally to each of the road components of that route. Thus from time dependant transition matrix $A_{30}$ we have a delay factor 0.55 for a road component $A_2_{150}$ and 0.55 for $A_7_{250}$. The reporting system then generates a weighted average delay factor for each road component using all the available data for different vehicles. The average is weighted according to the probability of the vehicles being on the road component (given by the probability of the route), and a decay factor in the case of earlier generated delay factors which would typically be of the order of a 10% linear reduction for each elapsed minute.

Thus the delay factors obtained for the vehicle by the above process would be as follows:

<table>
<thead>
<tr>
<th>Road Component</th>
<th>Delay Factor</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_2_{150}$</td>
<td>0.55</td>
<td>0.24</td>
</tr>
<tr>
<td>$A_7_{250}$</td>
<td>0.55</td>
<td>0.24</td>
</tr>
<tr>
<td>$A_1_{150}$</td>
<td>1.3</td>
<td>0.38</td>
</tr>
<tr>
<td>$A_1_{250}$</td>
<td>1.3</td>
<td>0.38</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

The system then looks at a particular road component ($A_1_{250}$) and looks at all the available data (for different vehicles) for that road component and calculates a weighted average as follows:

<table>
<thead>
<tr>
<th>Vehicle</th>
<th>Delay Factor</th>
<th>Decay Factor</th>
<th>Probability</th>
<th>Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.3</td>
<td>1.0</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>2</td>
<td>0.9</td>
<td>1.0</td>
<td>0.52</td>
<td>0.52</td>
</tr>
<tr>
<td>3</td>
<td>1.1</td>
<td>0.8</td>
<td>0.40</td>
<td>0.32</td>
</tr>
</tbody>
</table>

$$\text{WeightedAverage} = \frac{\sum (\text{delay factor} \times \text{weighting})}{\sum (\text{weighting})} = \frac{1.314}{1.22} = 1.077$$ (2)

The raw weighted average delay factor is then converted into one or more different forms of traffic report suitable for export to the outside world. A particular simple form would be a reporting of the grade of the delay, for example:
<table>
<thead>
<tr>
<th>Delay Factor</th>
<th>Grading</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 1.1</td>
<td>no delay</td>
</tr>
<tr>
<td>1.1 - 1.3</td>
<td>moderate delay</td>
</tr>
<tr>
<td>≥ 1.3</td>
<td>heavy delay</td>
</tr>
</tbody>
</table>

6 The Traffic Estimation Model

Another methodology to estimate road traffic parameters from location tracking data in cellular mobile networks (described in [8]) is possible when assumed that all entrances into and exits from the road are equipped with portals (e.g. toll stations), which can detect a passing vehicle, check the presence of an operating cellular phone on-board and identify it. In the case of presence of more than one phone on a specific vehicle, only one will be selected, disregarding the others. Identifications will be referred to the cellular network, which will provide the values $n_c^{(i)}(t_k)$ (the number of cellular phones on the vehicles in cell $i$ at the time instant $t_k$) and $v_c^{(i)}(t_k)$ (the number of cellular phones on the vehicles on the road that passed from cell $i$ to cell $i+1$ in the interval $(t_{k-1}, t_k]$) for all cells $i = 1, 2, ..., M$, at each instant $t_k$, $k=0,1,\ldots$. Let $s^{(i)}$ denote the road stretch that is included in cell $i$. Then, at every instant $t_k$, toll station will measure the values $c^{(i)}(t_k)$ (the number of vehicles that entered $s^{(i)}$ from outside the road in the time interval $(t_{k-1}, t_k]$) and $o^{(i)}(t_k)$ (the number of vehicles that left the road at $s^{(i)}$ in the time interval $(t_{k-1}, t_k]$). Moreover, let $L[km]$ be the length of the sub-stretch $s^{(i)}$ (supposed to be of equal length for simplicity); $n^{(i)}(t_k)$ the number of vehicles in $s^{(i)}$ at time instant $t_k$; $v^{(i)}(t_k)$ the number of vehicles in transit between $s^{(i-1)}$ and $s^{(i)}$ in the time interval $(t_{k-1}, t_k]$. Let us also suppose, that

$$n^{(1)}(t_k) = c^{(1)}(t_k)$$

(3)

So it means, that for the sake of simplicity we assume, that the number of vehicles in the part of the road, served by the first cell taken into account is equal to those data received from toll station, about the number of vehicles entered from outside the road. Having all these it is now possible to define the ratio vehicles/cellular phones in the first stretch as

$$r^{(1)}(t_k) = \frac{n^{(1)}(t_k)}{n_c^{(1)}(t_k)}$$

(4)

By letting $\bar{V}$ indicate the average free speed of the car on the road, and fixing $t_k = kT$, $k = 0, 1,\ldots$, with $T = L/\bar{V}$ (average time to travel a road stretch covered by a cell), one can define the ratio vehicles/cellulars in the cell $i-1$, as instant $t_{k-1}$, which has entered cell $i$ at the $t_k$, “updated” with the quantities of incoming and outgoing vehicles and with the measured number of mobile phones at $t_k$ as
\[ r_{\text{fast}}^{(i)}(t_k) = \frac{r^{(i-1)}(t_{k-1}) * n_c^{(i-1)}(t_{k-1}) + c^{(i)}(t_k) - o^{(i)}(t_k)}{n_c^{(i)}(t_k)} \]

One can also define

\[ r_{\text{slow}}^{(i)}(t_k) = \frac{r^{(i-1)}(t_{k-2}) * n_c^{(i-1)}(t_{k-2}) + c^{(i)}(t_k) - o^{(i)}(t_{k-1})}{n_c^{(i)}(t_k)} - \frac{r^{(i-1)}(t_{k-1}) * n_c^{(i-1)}(t_{k-1}) + c^{(i)}(t_k) - o^{(i)}(t_k)}{n_c^{(i)}(t_k)} \]

as the weighted average of the same ratio in the cell \( i - 1 \) at instants \( t_{k-2} \) and \( t_{k-1} \), also updated with entrances and exits in \( i \). In the fast stretches the ratios are propagated from one cell to another one every \( T \) seconds. In the slow one two ratios are supposed to cumulate. In practice we choose between two updating mechanisms as follows:

\[
r^{(i)}(t_k) = \begin{cases} 
  r_{\text{slow}}^{(i)}(t_k) & \frac{L}{V^{(i)}(t_{k-1})} < \frac{T}{2} \\
  r_{\text{fast}}^{(i)}(t_k) & \frac{L}{V^{(i)}(t_{k-1})} \geq \frac{T}{2}
\end{cases}
\]

So, we take slow one if its traversing time is less than a half the average time to traverse a generic stretch, fast otherwise. Now we can compute the density \( \rho^{(i)}(t_k) \) and the flow of cars \( q^{(i)}(t_k) \) in the part of the road \( s_i \) as

\[ \rho^{(i)}(t_k) = \frac{n_c^{(i)}(t_k)}{L}, i = 1, 2, ..., M \]

\[ q^{(i)}(t_k) = \frac{\nu_c^{(i)}(t_k)}{t_k - t_{k-1}} r^{(i)}(t_k) = \frac{\nu_c^{(i)}(t_k)}{T} r^{(i)}(t_k), i = 1, 2, ..., M \]

And now the velocity can be computed indirectly as:

\[ V^{(i)}(t_k) = \frac{q^{(i)}(t_k)}{\rho^{(i)}(t_k)} i = 1, 2, ..., M \]

Simulation results refer to a single direction of movement on a three-lane road with a single initial entrance and a final exit (i.e. no intermediate toll stations). This is the worst case condition because the presence of points that absorb or inject traffic improves the precision of the estimates. Evaluations have been performed on two road segments of 20 and 40 km respectively. Average error in estimations was about 3-5% and never exceeded 16%. Nevertheless the need of “toll-stations” on entrances to and exits from the road leaves practical implementation of the idea unclear.
7 Summary

As traffic telematics background shows loses due to the lack of available information are tremendous. Alone in year 1999 in Germany there were altogether 4.7 billion hours wasted in cars, 12 billion liters of fuel waste per year which gave 105 billion Euro of economic loses and its only in one country per one year. It is also known that in German road-network there is available only about 20 meters per car in case if all want to drive at the same time. That is because about 40 millions cars have to share hardly 800.000 km of roads. Therefore there is a tremendous need in customized travel information: accurate, actual, reliable which could be issued pre-trip, on-trip reporting accidents, traffic jams, traffic forecasts,... Those, based on the simple observation methods suffer from the accuracy and the complexity of the information integration received from untrusted sources. Fixed sensor deployment is expensive both to deploy and maintain. Those existing implementation from Tegaron (T-Traffic) made together with T-Mobile or Passo made with Vodafone include about 4000 sensors placed all over the Germany. Granularity and timeliness is very dependent on the intersensor spacing and cannot guarantee to provide trusted information at every place. Therefore the floating car data methods seem to be the most worth to consider.

Mobile communications are widespread in a large part of industrialized countries and cellular networks, by which mobile radio-communications are supported, can give directly or potentially a huge amount of frequently updated information on the position of their users. This information can be used to estimate on-line the traffic conditions of important roads and highways by exploiting the presence of mobile phones on boards of vehicles. The method described in the Traffic Estimation Model section seems to provide good results but practical implementation of so called “toll stations” which would control entrances to / exits from the roads and check for mobile phone presence is not very clear. In contrast to this, the first presented method, based on the probabilities seems to be easy to implement. Applied Generics has spent over two years developing a commercial road traffic information system [4], known as RoDIN24, based on the principles described above. Working with major networks in UK and monitoring equipment partners, the company has developed software enabling RoDIN24 to monitor a mobile operator’s network without affecting the network’s performance or capacity. Physically, RoDIN24 consists of a small network of standard Sun Solaris rack-mounted servers, deployed in the operator’s switching centres to minimize network connections. The results of the trial have been extremely encouraging.
References


