Acquisition of traffic and mobility data based on cellular network signalling

Daniel Elias[1*], Friedrich Nadler[1], Danilo Valerio[2], Pierdomenico Fiadino[2]

1. nast consulting ZT GmbH, Austria

2. FTW Forschungszentrum Telekommunikation Wien GmbH, Austria

* Lindengasse 38, 1070 Vienna, +43 1 523 47 33 28, elias@nast.at

Abstract

This paper describes various techniques that are developed for traffic analysis purposes based on cellular network signalling events. Starting from the mobility reported by the mobile devices of a cellular network, traffic volumes and origin-destination matrices are calculated, zones of interests are analysed and users on different modes of transport are identified for rural and urban areas and different times of the day (e.g. peak hours and night traffic). In addition a tool is described that can process that cellular data for import into expert traffic modelling software which assists during traffic planning and civil engineering exercises. The new extraction methods increase the accessibility of traffic data and reduce the costs of data collection in comparison to conventional collection and analysis methods. Furthermore the results are checked against reference traffic counts and additional traffic engineering statistics for validation and quality assurance purposes.

Keywords

cellular data, traffic modelling, origin destination relations, traffic analysis, zones of interest

Introduction

The research project VERMOBIL (an industrial research financed within the framework of the 2nd tender of the program ways2go of the research- and technology program IV2Splus) deals with the provision of data for traffic studies and traffic modelling purposes.

The data is generated out of cellular network signalling events, representing the mobility of mobile devices across cellular network spatial entities. In order to be accessible, mobile phones continuously inform the cellular core network about their current location. This data – properly anonymized to preserve citizens' privacy – can be collected and used for traffic planning purposes. The generated events can be analysed and aggregated in order to use them for different use cases in the field of traffic engineering.

Due to the vast amount of raw data that must be processed and the complexity of cellular network architecture and protocols, the analysis is performed via a multiple stage processing chain. First, a module identifies and filters anomalies from the data-set. Second, network mobility artefacts are excluded, e.g. changes of cells caused by fluctuations of the received signal strengths rather than actual mobility (ping-pong effect). Finally algorithms are deployed to capture the mobility of devices throughout the network, identify different modes of transport and aggregate the data for different planning and analysis purposes. In the context of this study, anonymized cellular network signalling events were provided by a major mobile phone operator in Austria.

The concept of using cellular data for extracting road traffic information is not new and has been the focus of several research projects in the past few years. Almost all previous studies consider mobile phone data related to "active" terminals. The position of these terminals, i.e. engaged in a call/data session, is known by the network at the cell level (or even with higher accuracy if signal strengths are considered). However, this approach suffers from two main limitations. First, the number of active terminals at any instant of time is several orders of magnitude smaller than the total number of terminals registered to the network. Thus, observing active terminals alone, produces a biased view of the overall human mobility. Second, obtaining information about cell changes and signal strengths might require to monitor several hundreds of network links (the so-called "A" and "Iub" interfaces). The costs of such a monitoring system would prevent a large-scale deployment and is sustainable only for small-scale field-tests. In contrast to previous research, the study exploits the more complete signaling data captured between the Radio Access Network (RAN) and the Core Network (CN) of a cellular operator. The position of all "idle" terminals country-wide can be observed (at a coarser spatial granularity) by monitoring few network links. Since idle terminals are the overwhelming majority of the mobile terminal population, this approach allows to reach much better coverage, at lower costs.

Techniques and use cases

Identification of trajectories

Based on the definition of different trajectories (e.g. along a motorway, regional, rural or urban road) an algorithm assesses the cellular network events and aggregates the number of devices that are identified along the defined trajectory. The query can include a strict sequence of cells or can also allow skipping of some of the defined locations in case a user takes a detour or brake along the trajectory.

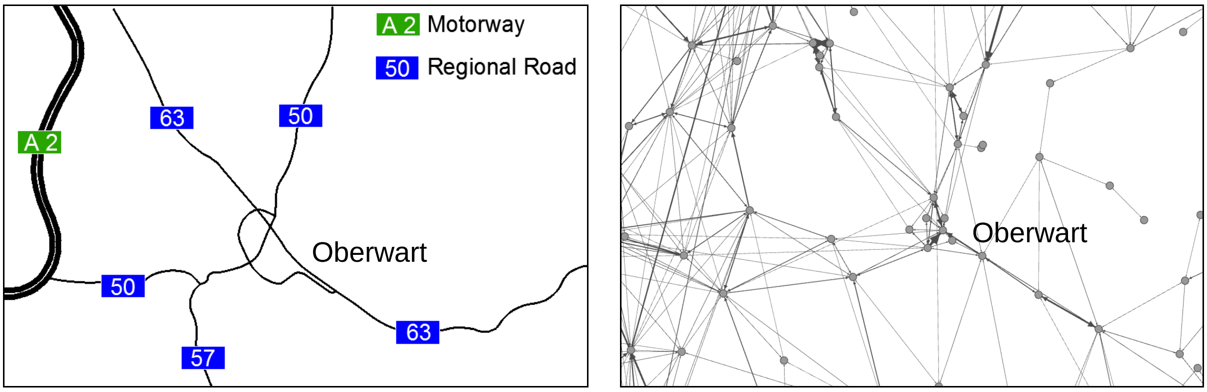


Figure - Example for analysis of frequent connections between neighbouring cells

Identification of different modes of transport

Depending on the network configuration various techniques can be used to identify and different modes of transport. Some of the methods that were developed during the study include the analysis of travel times of individual users, the reverse projection of location area updates based on train time tables, analysis of specific routes that can only be accessed through the subway train network. Furthermore peaks of e.g. location area updates can be analysed and disaggregated to identify users on different modes of transport.

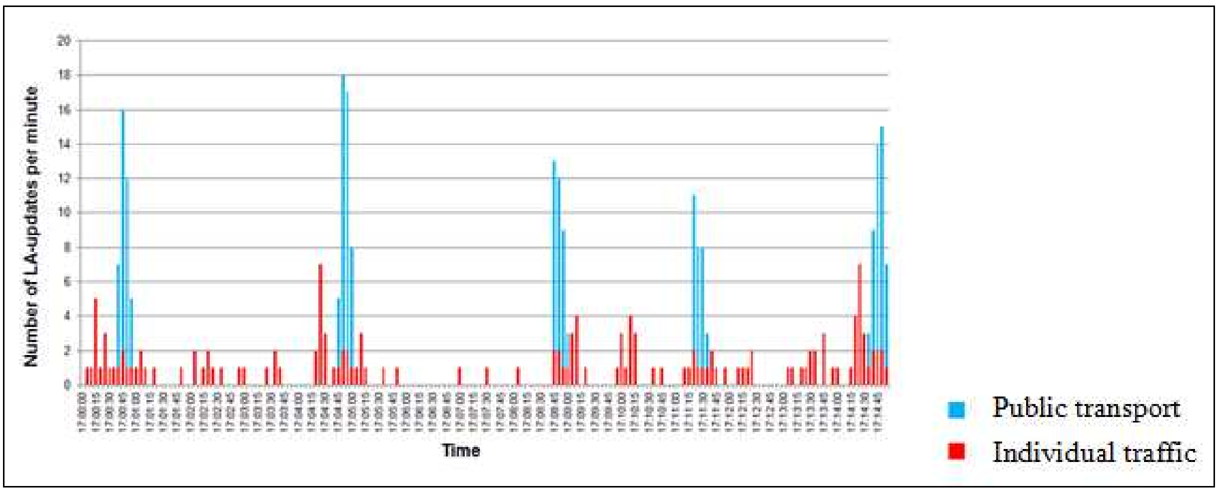
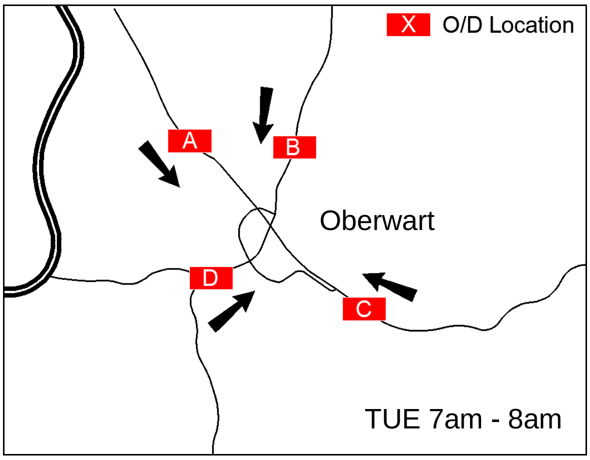


Figure - Example for peak time disaggregation

Provision of origin-destination matrices

Algorithms are developed that focus on the extraction of origin-destination relations. Based on a set of cells that act as origin/destination locations, the number of users between those defined points is aggregated on a certain level e.g. daily, hourly traffic. In addition the user trajectory algorithm supports the estimation of the share of different road corridors between origin and destination points (e.g. estimation of the total daily vehicle traffic in an administrative unit such as the region Oberwart in Burgenland). The methods that are developed also consider dwelling times at certain locations as well as disturbing ping-pong effects.



Location	Incoming traffic	Percentage total
A	435	19%
B	397	17%
C	669	29%
D	801	35%
Total	2302	100%

Figure - Example for commuting analysis

Analysis of zones of interest

In addition to the preparation of origin-destination matrices the methods can also assist during the evaluation of zones of interests for e.g. shopping centres, health facilities or event locations. The developed algorithms can be used to estimate the share of incoming traffic from each direction and can assist during evaluation of commercial advertisement along specific road corridors.

Automated processing for traffic modelling tools

The algorithms that are developed as part of the study are based on individual cell-id definition which can be done either automatically or manually. The manual definition is assisted through visualization of available cells and their corresponding details in Google Earth. Google Earth in conjunction with a geographical information system can also be used to define geographical origin-destination points, points of interest or even parts of the road network which can be processed for import into traffic modelling tools. Parallel, the results of the user trajectory analysis for a specific region (e.g. Oberwart) are also recoded for import into the traffic modelling tool. Consequently, using the same identification number of the defined points and roads, the modelling tool is able to visualize the results of the trajectory analysis (static link chains or dynamic traffic reallocation). The processing concept includes basics for data encoding as well as fundamentals for the definition of the traffic graph that is needed to process the results.

Validation and quality assurance

The results of the automated data processing only include a certain percentage of the total traffic which can further contain different modes of transport. In order to ensure a high data quality with low interference (different modes of transport, parallel road sections) the results are checked against reference traffic counts, average travel times and historical traffic reports. Furthermore the expertise that is developed within the study can also include different data sources.

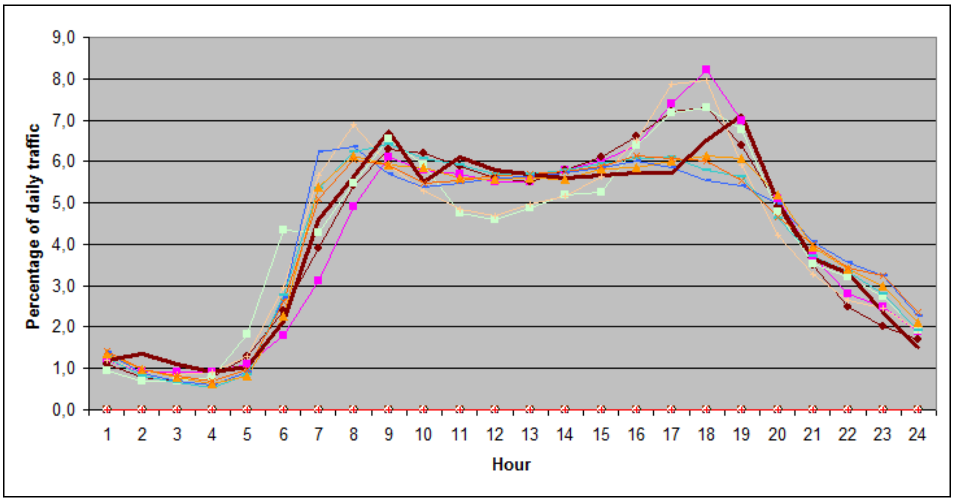


Figure - Example of comparison of traffic estimation with reference traffic counts

Conclusion

The definition of different cellular network event processing tools gives access to a wide variety of use cases in the field of traffic engineering and commercial analysis. Due to the convertibility of the input variables it is possible to conduct analyses for small areas as well as larger regions such as the whole Austrian area. The data extraction methods result in an increased availability and quality of traffic data for planning purposes as well as telematic and modelling applications. Furthermore costs for data collection can be reduced.

Acknowledgements

This work is the result of the research cooperation between nast consulting ZT GmbH and FTW (Forschungszentrum Telekommunikation Wien GmbH). Members of both organisations have spent an extensive amount of time to ensure a high quality standard of results.

References

1. Project VERMOBIL: research cooperation between nast consulting ZT GmbH and FTW (Forschungszentrum Telekommunikation Wien GmbH) (2010-2012). Industrial research financed within the framework of the 2nd tender of the program ways2go of the research- and technology program IV2Splus