How Mobile Phone Handovers reflect Urban Mobility: A Simulation Study

Thierry Derrmann, Raphaël Frank, Thomas Engel
Interdisciplinary Centre for Security, Reliability and Trust
University of Luxembourg, UL
Belval, Luxembourg
thierry.derrmann@uni.lu, raphael.frank@uni.lu,
thomas.engel@uni.lu

Francesco Viti
Faculty of Science, Technology and Communication
University of Luxembourg, UL
Luxembourg City, Luxembourg
francesco.viti@uni.lu

***Abstract*— We propose a novel way of estimating Macroscopic Fundamental Diagrams (MFD) (or often also called Network Fundamental Diagrams) from mobile phone signaling data under the assumption that vehicles can be identified from the data stream. We run a simulation study to identify whether MFDs can be constructed from this type of data. We co-simulate the road traffic in Luxembourg City and one mobile operator's LTE network user plane, and show that mobile network base station clusters cover road network partitions of coherent behavior. Our results indicate that the relationships between handovers (flow) and attached phones (accumulation) in these base station clusters constitute MFDs. We validate our results by comparing the road network MFDs to the ones obtained from the mobile network.**

Keywords—macroscopic fundamental diagram; mobile network; cellular data; handover;

I. Introduction

Road traffic management is a complex task requiring detailed insight on mobility and traffic metrics. In this context, road traffic engineers rely on predictive models for different road categories and settings. In the urban context, the complexity of the underlying traffic patterns, and the dynamics of congestion make often render predictions inaccurate. On the other hand, regularities in the dynamics of network states have been observed through network-wide data, in early studies by Godfrey [28] and later Mahmassani and Williams [29, 30]. Extending upon these ideas, in particular, urban networks can be described by the so-called Macroscopic Fundamental Diagrams (MFD) as named and theorized by Geroliminis and Daganzo [1]. MFDs serve to describe the flow, density and velocity relationships of the cumulated data from multiple detector sources within a partition of the network, where traffic states and road capacity are relatively homogeneous. These models therefore allow the estimation of traffic conditions for partitions of the network and enable traffic management techniques such as perimeter control or coordinated intersection control [12, 13].

MFDs are synthetic but powerful metrics that quantify and explain the interaction between road capacity, travel and driving behavior-related parameters such as routing/rerouting, as well as characteristic vehicle speeds and car following behavior. It postulates that if a sufficiently large amount of data about traffic states in a network is collected, and the (sub) network topology has a sufficient level of regularity in terms of route flow distribution, then state variables such as vehicle density and the total network throughput are clearly related by a concave function. This function expresses the transition between uncongested conditions to congested states, characterized in urban systems by frequent conditions of queue blocking and gridlock phenomena. Theoretical and empirical studies contributing to gain insight into the properties of MFD focused on deriving relations starting from analytical and simulation-based Dynamic Traffic Assignment theory [21], on assessing the impact of traffic control [22], and on capturing hysteresis phenomena in congested networks [23].

On the practical side, obtaining data that exhibits MFD characteristics is difficult, and the traffic flow theory community has recently launched a challenge to compile a suitable empirical data for MFD modeling [14]. In this context, there is a potential use for mobile phone data. Mobile network operators can provide abundant data as the penetration rate and connectivity of mobile phones is ever-increasing.

Different types of mobile network data have been and are being made available by mobile network operators in different countries, e.g. in research challenges such as the Orange Data4Development and the Telecom Italia Big Data challenges. Among these datasets are *Call Detail Records* (CDR), which provide counts of calls, data and text messages sent to and from different mobile base stations.

Another type of mobile network data used for analysis is *signaling data*. One kind of signaling data is that of handovers, i.e. how often and when users change between base stations during active connections. With the recent and future radio technologies such as UMTS, LTE and the upcoming 5G, the coverage areas of these base stations keep becoming smaller. Thus, the spatial resolution of the mobile phone data increases, yielding more valuable information for traffic applications. Further, it is possible to use this type of data in real time, and the temporal resolution of the data aggregation is typically adaptable.

In the field of transportation, mobile phone data have mostly been used for predicting highway travel times [2, 3], and for demand estimation [4]. In this paper, we propose to use mobile phone signaling data for a novel application in transportation: we use mobile phone data to estimate Macroscopic Fundamental Diagrams of network partitions.

The main motivation for using mobile phone data instead of other sensor types such as loop detectors lies in the high coverage, as it is potentially available anywhere where there are mobile networks. Also, mobile network operators optimize their network to offer high service quality among the most frequented locations in their network, which means that the infrastructure typically matches human mobility and higher density of base stations in the urban areas. The main drawback is the potential need for data preprocessing to identify vehicles. However, this does not prevent real-world applicability of our results. It is possible to identify connected vehicles by their Mobile Equipment Identifiers (IMEI number), and base the analysis on these equipments only. This will be increasingly usable in the coming years, as today's new vehicles make use of mobile data services and are routinely equipped with SIM-Cards, e.g. for the European eCall system. In the light of privacy concerns, there is also the possibility of leveraging cell dwell times, i.e. the time between consecutive handovers. Janecek et al. [2] have demonstrated that filtering out vehicles from signaling data: On highways, they are typically among the fastest moving mobile phones or devices (called "User Equipments" in LTE), and can thus be identified by their dwell times, i.e. the time between two consecutive handovers.

However, identifying vehicles among users may be unnecessary in practical implementations, if it is possible to extract MFDs even in the presence of 'noise' caused by stationary users. The study of this effect is subject of future analysis on real data, and not in the scope of this work. For this study, we assume that vehicles can be identified from the signaling data stream.

The kind of data needed for this type of analyses is twofold, as we need measures proportional to vehicle flow and density. In [5], Geroliminis and Daganzo show how the accumulation of vehicles within a region and their outflow from it, construct a macroscopic fundamental diagram. Thus, we propose to apply this concept to mobile phone signaling data, where the flows can be estimated from the number of handovers, and vehicle densities can be estimated from the amount of phones associated with a base station.

II. SIMULATION AND METHODOLOGY

Motivation

The main motivation of this work to evaluate whether the estimation of MFDs from mobile phone signaling data is principally possible. To the best of our knowledge, such an approach has never been evaluated.

Our research hypothesis is that the mobile network and road network behavior are correlated. More precisely, that the counts of handovers and connected users per base station can be used as proxy metrics for road network flows and densities, respectively. As mobile network operators aim to fully cover densely populated areas and the main corridors of mobility, we believe it is sensible to evaluate their aptitude of capturing population dynamics.

To verify whether these ideas hold up in an ideal case, we want to use simulation to generate purely vehicular signals, i.e. data from drivers (or connected vehicles) only. If this data exhibits MFD behavior, it will open the way for research on how to extract road traffic MFDs from real mobile network signaling data by adequately filtering and preprocessing the data. We will go into the practical implications of the restriction to vehicle-only data in the discussion section, and propose some possibilities for generalizing to full population data sets.

Road and Communication Network Co-Simulation

The general goal is to collect vehicular in- and outflows and the density of LTE base stations (eNodeBs) from the simulation. As stated, we only simulate vehicles, not pedestrians or stationary users.

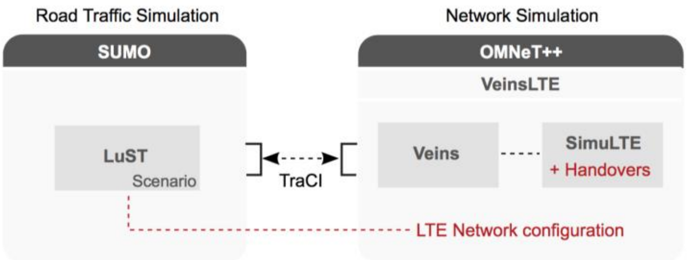

Fig. 1: Overview of the LuST-LTE simulation package

Scenario

The simulation scenario we base our study on is the LuST scenario by Codecà et al. [15] for the microscopic traffic simulator SUMO [8]. The scenario provides 24 hours of mobility in the Luxembourg City area (155 km^2). Note, however, that the results of this study are based on a half-day simulation (5:00-14:00) run, as this exhibits the different traffic states we need. We ran the scenario with 10% re-routing probability of vehicles, a deliberately low value to observe congestion in the network, as many vehicles take similar instead of alternative routes.

The VeinsLTE framework by Sommer, Hagenauer et al. [6, 7], realizes the connection between the microscopic road traffic simulation SUMO and the communication network simulator OMNeT++ [9], simulating the LTE connectivity of cars/drivers.

Through VeinsLTE, the SUMO simulation is synchronised with the OMNeT++ communication simulator [9] to reflect the mobility of vehicles inside the mobile phone network. That means that SUMO simulates the movement of vehicles and OMNeT++ computes signal strengths, connectivity and communication of the moving vehicles. More precisely, the LTE network is simulated using the SimuLTE library [10], to which we added an implementation of handovers based on Signal-to-Interference-plus-Noise Ratio (SINR). On the other side, SUMO provides the traffic demand and microscopic modelling and we use it to generate Floating Car Data for validation purposes. The identifiers of vehicles are matching between both simulators and thus allow e.g. to map signal strengths by routes taken.

Figure 1 shows the interaction between the different components. The components in red were added to the existing frameworks. The LTE network configuration consists in a mapping of 72 eNodeBs (LTE base stations) to the simulation coordinates. The original coordinates of the eNodeBs were estimated using the OpenCellID API. We took the assumption that a single base stations hosts about 7 cells, which corresponds to an urban LTE deployment by Vodafone Germany [26].

We named the simulation package 'LuST-LTE' and it is available online [25]. More details regarding the simulation package can be found in [27]. The downloadable package features all the necessary components to run the simulation used in this paper. We also plan to make the result dataset available on the repository [25].

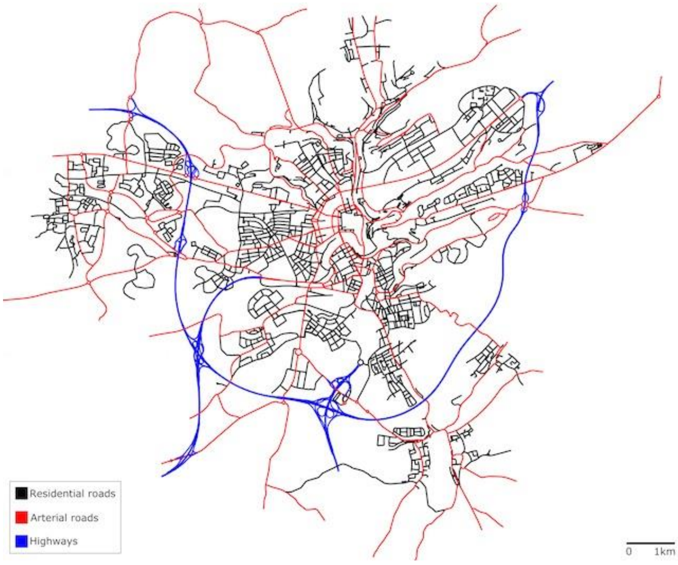

Fig. 2: The LuST scenario road network [15]

Signaling Data

Automotive security company Giesecke & Devrient (G&D), expects that approximately 75% of all cars shipped in 2020 will be delivered with integrated connectivity, and according to Scotiabank, in 2014 approximately 10% of cars included connectivity [31].

A study on MFDs from Floating Car Data by Gayah et al. [26] confirmed low penetration rates at around 5% equipped cars to be sufficient for error percentages below 10%.

We chose a conservative penetration rate of 3% for our simulative study, accounting for the "missing" terminals in real data that are disconnected. Note that in a real mobile network, passive devices (without an active connection) are only traceable to location area scale, i.e. multiple base stations of spatial resolution. Thus, we opted to simulate only the proportion of users with an active data or call connection, for which the currently visited cell is known to the mobile network operator.

The simulated signaling data that we generate consists of handover data (timestamp, previous and new cell-ids), and the amount of users associated with eNodeBs in 10 minute intervals. We chose this interval size according to data export resolutions available in real systems.

Partitioning the mobile and road networks

In order to obtain useful clusters for MFD estimation, we based ourselves on the findings of Geroliminis et al. [1], who stated that the MFD phenomenon emerges in areas larger than 10 km^2. As the inner city surface of Luxembourg City is ~40 km^2, we opted for 4 clusters.

Using our domain knowledge of Luxembourg City's topology, we split the road network into 4 partitions, as shown in Fig. 3:

- City Center (red)
- Hollerich (blue)
- Bonnevoie-Cents (green)
- Kirchberg plateau (violet)

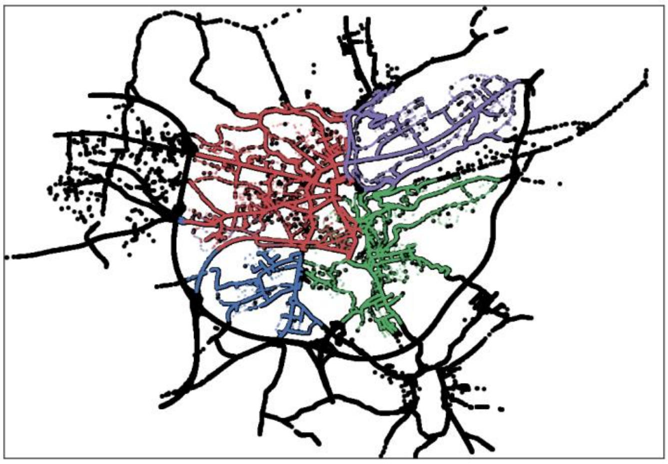

Fig. 3: Clustering of the Luxembourg City urban road network

From our dataset, we know which roads map to which eNodeB, so we can use the road network partition the mobile network accordingly. This way, we can compare statistics on the mobile network to those from the road network, in particular flows and accumulation of both networks.

Mobile phone MFD estimation

In order to estimate the MFD of a mobile network partition, we evaluate the handovers leaving the partition and the accumulation of vehicles within the partition:

$$Outflow(C,t_{start},t_{end}) = avg(n_{handovers}(t_{start},t_{end},s,d) \mid s \in C, d \in C) \quad (1)$$

$$Accumulation(C,t_{start},t_{end}) = avg(\sum n_{connected}(t_{start},t_{end},s)) \quad (2)$$

III. Results

Figure 4 shows the number of handovers per 10 minute window inside the simulation. We can see the loading period in the morning peak around 8:00 and the recovery period after 10:00.

The mobile network handover counts roughly double during the peak hour, reflect the increased demand and flows in the road network. Note that due to the low re-routing probability employed in SUMO (only 10%), the congestion observed is severe, and we observe grid-lock. Hence, the road network does not recover quickly. However, this is useful to observe the full traffic pattern spectrum that we need to fully characterize the area MFDs.

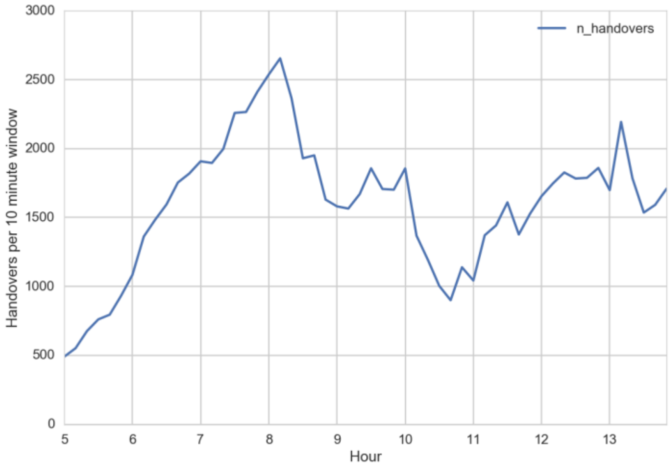

Fig. 4: Simulated mobile network handovers per 10 minute window

Figure 5 shows the main results of the study. On each of the graphs, each data-point represents a 10-minute window. The left-side plot shows the 'classic' MFD representation of outflows vs. accumulation, but estimated from the mobile phone data according to Equations 1 and 2.

The center graphs show the 'real' velocity-accumulation relation, and the right hand side graphs shows the equivalent using the mobile phone data (by evaluating the ratio between outflows and accumulation).

The center plots show very smooth velocity-accumulation profiles, with low variance.We can see that the mobile phone data graphs exhibit larger variances, but follow the same trend with respect to accumulation.

Thus, we also computed the Pearson correlation coefficient between the mobile network velocity proxy and the actual space-mean velocity for each road network partition. Table 1 gives the correlation coefficients by cluster, which vary between .81 and .95 and are thus strong predictors. In order to predict velocities from the mobile network data, it is then sufficient to multiply the proxy by a scale factor that is dependent on the number of exit lanes from the partitions and the length of road segments covered by the mobile network base stations. In practice, the scaling can be estimated using Floating Car Data.

Partition	**Pearson: $v^{true} \sim v^{proxy}$**
City Center (red)	0.95
Hollerich (blue)	0.81
Bonnevoie-Cents (green)	0.9
Kirchberg plateau (violet)	0.85

Table 1: Partition correlations: simulated ground truth vs. approximated velocities

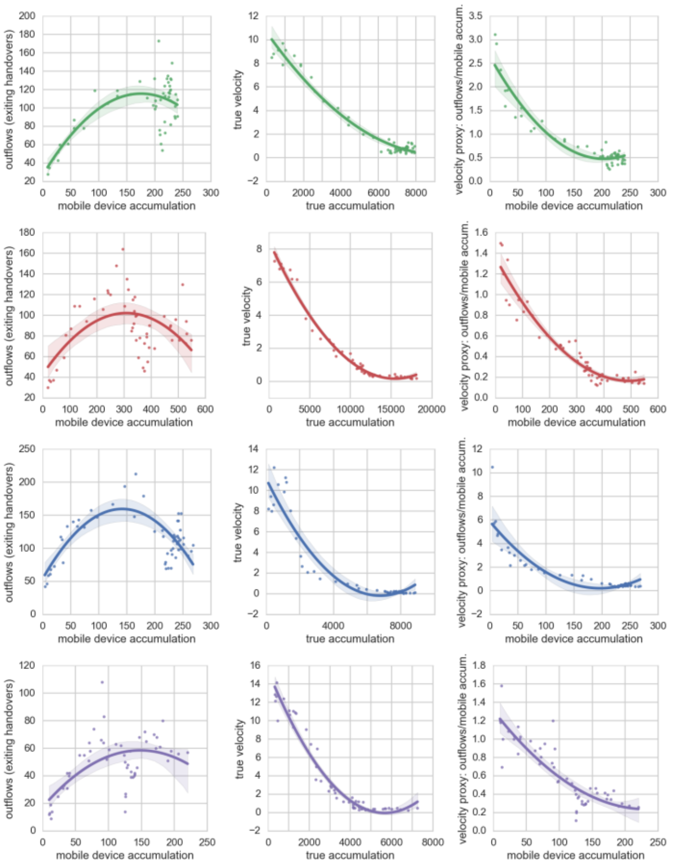

Fig. 5: Results for the 4 partitions (colors match Fig. 3)

Overall, the typical MFD profiles that are known from loop detector data are also observed with the mobile phone data, but with more variance.

We can also see that all partitions' outflow-accumulation graphs exhibit some deviation from the mean in the middle or high accumulation data range. Looking at these cases more closely, we found that flows fall below the mean during the de-loading period (~10:30-12:00), and exceed it during the peak hour (~7:30-8:30).

The results show that the simulated mobile network MFDs indeed reflect the actual traffic in the road network, and that mobile network data has the potential to serve as a novel way of estimating MFDs.

CONCLUSION

In this paper, we have shown that mobile network clusters exhibit MFDs if vehicles can be identified from the mobile network users. Using the simulated handover information, we observed Pearson r-values between 0.8 and 0.95 for the correlation between actual road velocities and the mobile network MFD estimates. Thus, our analysis shows that exiting handover counts of a partition are a good proxy for flows at a penetration rate of 3% of connected mobile phones in traffic.

Intuitively, it makes sense that vehicular mobility would make up a large portion of overall handover counts, which was one of the main motivating aspects of this work. This means that the passive, unconnected part of the user-base is not problematic for the MFD estimation. This is an important, but surmountable practical restriction. The extraction of vehicles from the signaling data requires preprocessing steps at the mobile network operator level. More concretely, techniques as proposed by Janecek et al. [2] allow the identification of drivers' phones by their dwell times between base station pairs. Alternatively, it can be sufficient to measure the number of calls, data and SMS activities within cells as a proxy metric for vehicular accumulation. This would yield biased accumulation values, from which the vehicular traffic could be estimated by regression against standard datasets, e.g. loop detector or census data.

Following up on the promising results of this study, we are currently working on a data set of mobile network to confirm this paper's findings in the field, with the additional constraints that this entails: stationary users, 'ping-pong' handovers and noisy traffic data. For this purpose, future work consists in an optimized partitioning of the mobile phone network in order to obtain low-variance, homogeneous partitions of the underlying road network. Such a partitioning technique will especially help with real handover data, where there are stationary users and small effects that can lead to noisy data. In this case, variance needs to be reduced to a minimum in order to enable valid predictions of traffic states.

ACKNOWLEDGMENT

The authors would like to thank the National Research Fund of Luxembourg (FNR) for providing financial support through the CORE 2013 MAMBA project (C13/IS/5825301).

REFERENCES

[1] N. Geroliminis, C. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings." *Transportation Research Part B: Methodological* 42.9 (2008): 759-770.

[2] A. Janecek et al., "The Cellular Network as a Sensor: From Mobile Phone Data to Real-Time Road Traffic Monitoring." *IEEE Transactions on Intelligent Transportation Systems* 16.5 (2015): 2551-2572.

[3] H. Bar-Gera, "Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel." *Transportation Research Part C: Emerging Technologies* 15.6 (2007): 380-391.

[4] S. Di Donna, G. Cantelmo, and F. Viti, "A Markov chain dynamic model for trip generation and distribution based on CDR." *Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2015 International*

Conference on. IEEE, 2015.

[5] N. Geroliminis, Nikolas, C. Daganzo, "Macroscopic modeling of traffic in cities." *TRB 86th annual meeting*. No. 07-0413. 2007.

[6] C. Sommer, R. German, and F. Dressler, "Bidirectionally coupled network and road traffic simulation for improved IVC analysis." *IEEE Transactions on Mobile Computing* 10.1 (2011): 3-15.

[7] F. Hagenauer, F. Dressler, C. Sommer, "Poster: A simulator for heterogeneous vehicular networks." *2014 IEEE Vehicular Networking Conference (VNC)*. IEEE, 2014.

[8] M. Behrisch, et al., "SUMO–simulation of urban mobility: an overview." *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation*. ThinkMind, 2011.

[9] A. Varga, "The OMNeT++ discrete event simulation system." *Proceedings of the European simulation multiconference (ESM'2001)*. Vol. 9. No. S 185. sn, 2001.

[10] A. Virdis, G. Stea, and G. Nardini, "SimuLTE-A modular system-level simulator for LTE/LTE-A networks based on OMNeT++." *Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH), 2014 International Conference on*. IEEE, 2014.

[11] L. Codecà, R. Frank, and T. Engel, "Luxembourg SUMO traffic (LuST) scenario: 24 hours of mobility for vehicular networking research." *Vehicular Networking Conference (VNC), 2015 IEEE*. IEEE, 2015.

[12] J. Haddad and N. Geroliminis, "On the stability of traffic perimeter control in two-region urban cities." *Transportation Research Part B*, 46(9):1159–1176, 2012.

[13] J. Haddad, M. Ramezani, and N. Geroliminis, "Cooperative traffic control of a mixed network with two urban regions and a freeway." *Transportation Research Part B*, 54:17–36, 2013.

[14] J. Laval, "MFD Dataquest." *MFD Dataquest Website*. Subcommittee of AHB45 TRB Committee, <https://sites.google.com/a/jltraffic.com/mfd-dataquest/home>.

[15] L. Codecà, R. Frank. "LuST Scenario." *VehicularLab Website*. VehicularLab, 29 July 2016, <http://www.vehicularlab.uni.lu/lust-scenario>.

[21] H.S. Mahmassani, M. Saberi, A. Zockaie, "Urban Network Gridlock: Theory, Characteristics and Dynamics". *Transportation Research Part C: Emerging Technologies* 36: 480-496.

[22] V. V. Gayah, X. Gao, A. Nagle, "On the impacts of locally adaptive signal control on urban network stability and the Macroscopic Fundamental Diagram". *Transportation Research Part B: Methodological* 70: 255-268.

[23] M. Saberi, H.S. Mahmassani, "Hysteresis and Capacity Drop Phenomena in Freeway Networks: Empirical Characterization and Interpretation. *Transportation Research Record* 2391: http://dx.doi.org/10.3141/2391-05

[24] "OpenCellID API." *OpenCellID Website*, 29 July 2016, <http://www.opencellid.org>.

[25] T. Derrmann, L. Codecà, R. Frank. "LuST-LTE Simulation Package." *VehicularLab Website*. VehicularLab, 29 July 2016, <http://www.vehicularlab.uni.lu/lust-lte>.

[26] V. Gayah and V. Dixit. "Using Mobile Probe Data and the Macroscopic Fundamental Diagram to Estimate Network Densities - Tests Using Microsimulation." *Transportation Research Record: Journal of the Transportation Research Board, No. 2390, pp. 76–86.*

[27] T. Derrmann, S.Faye, R. Frank., T. Engel "Poster: LuST-LTE: A Simulation Package for Pervasive Vehicular Connectivity." *Vehicular Networking Conference (VNC), 2016 IEEE*. IEEE, 2016.

[28] J.W. Godfrey, "The mechanism of a road network", Traffic Engineering and Control, Vol. 11 (7), 323-327.

[29] H.S. Mahmassani, J.C. Williams, R. Herman, "Performance of urban traffic networks". Proceedings of the 10th International Symposium on Transportation and Traffic Theory, Elsevier Science Publishing, 1-20, 1987.

[30] J.C. Williams, H.S. Mahmassani, R. Herman, "Urban traffic network flow models", Transportation Research Record: Journal of the Transportation Research Board, No. 1112, 78-8, 1987

[31] " Giesecke & Devrient Expects 75% of all Cars to Be Shipped with Integrated Connectivity by 2020 API." *Giesecke & Devrient Website*, 15 February 2016,
<https://www.gi-de.com/en/about_g_d/press/press_releases/global_press_release_39873.jsp>