Improving the Accuracy of IVC Simulation using Crowd-sourced Geodata

Christoph Sommer, David Eckhoff and Falko Dressler
Computer Networks and Communication Systems, Dept. of Computer Science, University of Erlangen, Germany
{csommer,eckhoff,dressler}@cs.fau.de

***Abstract*—We discuss the use of crowd-sourced geodata in simulative evaluations of Inter-Vehicle Communication (IVC) protocol designs. Typically, network simulation tools, which have been improved over decades of network research, are used for evaluating communication systems. In the area of IVC, however, additional challenges have to be met. Most important, the mobility of vehicles in network simulation needs to be represented accurately, e.g., using road traffic microsimulation techniques. These can be integrated with network simulation tools in order to provide a holistic view on the overall system performance. Obviously, the quality of these approaches inherently depends on the quality of provided map data. The OpenStreetMap project provides a community-maintained repository under an open license model. The available crowd-sourced geodata not only consists of road topology data but also includes fine-grained details such as traffic lights, speed limits, and even information about buildings, which represent obstacles for wireless communication. Using our Veins simulation framework, we show that this data increases the accuracy of IVC simulation.**

I. INTRODUCTION

Networks of communicating cars are at the center of the challenging research field commonly termed Inter-Vehicle Communication (IVC) [1]. Wireless communication techniques provide the basis for this communication – most recently, the IEEE 802.11p standard has been defined [2]. Network layer requirements are extremely dependent on the application scenarios, ranging from infrastructure-based services and Vehicular Ad Hoc Network (VANET) routing to broadcast-based information dissemination.

Broadly speaking, two major domains of IVC research can be identified [3]: First, *safety applications* aim at mitigating or avoiding danger to their users. This goal is achieved by exchanging emergency warnings or providing intersection coordination and collision avoidance. Secondly, *comfort applications* help improving the passenger comfort and traffic efficiency and/or optimizing the route to the destination. Exchanged data includes traffic information, weather and road conditions, and others. The development of an efficient Traffic Information Systems (TIS) is among the most frequently cited applications in this field [4]. The primary objective of TIS is to provide traffic relevant information to the driver of a vehicle in various situations. Supplementing the information available from pre-installed systems (e.g., the navigation system including maps, and local sensors such as GPS) wireless communication is used to acquire additional data from remote systems.

Complementary to testbed experiments that are often limited in scale and variety, simulation is one of the most important methods for performance evaluation of IVC protocols. The quality of the simulation experiments depends on a number of factors. It has long been established that a detailed network traffic model is not enough and an accurate mobility model, i.e. a road traffic model, is needed [5], [6].

The most obvious of causes is that the impact a TIS has on the behavior of drivers (and, hence, on road traffic) is heavily constrained by road topology, e.g. the number and relative weight of possible routes, traffic lights, and access or turn restrictions. Moreover, though, the road traffic model also affects the simulated network traffic. The most prominent way is by means of speed limits and lane counts, which influence a node's interconnection time with surrounding nodes and the number of neighbors a node has.

If the goal of simulations is a realistic simulation model, using an accurate city map means that network researchers can focus on their particular domain of expertise with little to no consultation from traffic modeling experts. Yet, if the degree of realism is high enough, such simulations could ultimately also be taken as a basis for predictions in specific application scenarios, e.g. traffic flow optimizations for a specific city.

In this paper, we study challenges in obtaining accurate and meaningful results from IVC simulations, with a special focus on Traffic Information Systems. In particular, we study crowd-sourced geodata from the OpenStreetMap project as a means for improving the realism of simulation experiments. Besides road information, especially annotation of speed limits, traffic lights, and even buildings suggest using this kind of geodata in road traffic simulations.

The rest of the paper is structured as follows. In Section II, we briefly introduce concepts of IVC protocol simulation. We then outline the capabilities of OpenStreetMap data, which is freely available for public use in Section III. The impact of using this data is discussed in Section IV. Finally, Section V concludes the paper.

II. IVC PROTOCOL SIMULATION

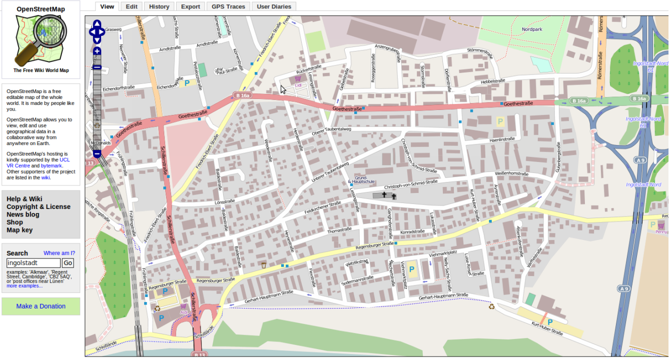
In the research community there exists a number of simulation frameworks for network simulation. Developed over several decades, the existing tools are quite accurate and well-established in the community. Typical examples include open source solutions such as ns-3, OMNeT++, and JiST/SWANS. However, none of these frameworks has been designed for simulating vehicular networks, in particular focusing on the mobility models.

In order to extend this functionality, solutions such as ASH [7] for JiST/SWANS and a vanet-highway model for ns-3 have been developed, which are using IDM/MOBIL [8] for modeling mobility on a linear segment of road. However, while these frameworks might excel in the simulation domain they were designed for, i.e. communication among freeway traffic, they cannot easily be adapted for modeling the effect a traffic information system might have on traffic flows in a two-dimensional road network, e.g., influencing traffic light timings or a driver's choice of route. Other frameworks, like GrooveNet (Roadnav, [9]) and SWANS++ (STRAW [10]) solve this problem by relying on mobility models designed for two-dimensional mobility. Still, they offer either no possibility for traffic information system behavior to influence road traffic, or they rely on geographically exact but overly simplistic map data representation (most commonly based on the U.S. Census Bureau's TIGER[1] data sets) for simulations, which lack most of the key road attributes that are pivotal for accurate modeling of road traffic flows.

In contrast, the required level of detail in mobility simulation has always been offered by dedicated road traffic microsimulation environments, which have been developed for this specific purpose and by experts from the domain of traffic engineering, most notably the SUMO [11] simulation environment. Such microscopic road traffic simulators (as opposed to mesoscopic or macroscopic approaches) model the behavior of single vehicles and interactions between them. Thus, researchers from the domain of network simulation took the logical step to integrate (or couple) simulators from both domains and form a simulation framework that can do both – accurately model network traffic and simulate realistic road traffic [12].

Two exemplary toolkits that follow exactly this approach are TraNS [13] and our Veins[2] [14] simulation framework, which integrate SUMO with the network simulators ns-2 [15] and OMNeT++ [16], respectively. Both toolkits rely on the TraCI interface [17] for coupling the different simulation environments. This allows the road traffic microsimulation to be used as a mobility model in the network simulation component. More important, it allows events in the network simulation to influence (via a human driver behavior model [18]) the course of the road traffic microsimulation.

With the availability of these simulation frameworks the accuracy of IVC simulations thus hinges on the degree of realism with which the simulation scenario was modeled. This, in turn, requires access to geodata of sufficient degree of detail. Because of the amount of work that assembling this body of data requires, we advocate the use of crowd-sourced geodata for this task. One possible solution, and the one we chose, is to employ an extension of SUMO to base scenarios on OpenStreetMap data for simulating the microscopic mobility of vehicles. We discuss details of the OpenStreetMap project and its capabilities in the following section.



OpenStreetMap is a free editable map of the whole world. It is made by people like you.

OpenStreetMap allows you to view, edit and use geographical data in a collaborative way from anywhere on Earth.

OpenStreetMap's hosting is kindly supported by the UCL VR Centre and bytemark. Other supporters of the project are listed in the wiki.

examples: 'Alkmaar', 'Regent Street, Cambridge', 'CB2 5AQ', or 'post offices near Lünen' more examples...

Fig. 1. Section of the city of Ingolstadt, Germany as displayed on the OpenStreetMap website

III. USING OPENSTREETMAP DATA

The OpenStreetMap project collects and provides user-generated geodata under the Creative Commons Attribution-ShareAlike 2.0 license. Exported data is thus available for both commercial and non-commercial use. As OpenStreetMap is a collaborative project that started in late 2004, area coverage, level of detail, and accuracy of data are all heavily dependent on the number of contributing users. Therefore, while overall geodata quality is rising steadily, care still needs to be taken when exporting data for use in simulations.

The OpenStreetMap web frontend,[3] reproduced in Figure 1, is capable of displaying this geodata in the form of a detailed 2D map, including, for example, streets, buildings, parks, railroads, and rivers. It also allows users to edit and export geodata. There are no limits on what attributes can be entered into the database, meaning that users can define new types of areas, however, only a set of predefined attributes will result in a distinct style of visualization.

The envisioned workflow of OpenStreetMap is as follows: A GPS data logger is used to create and upload traces to the database, which supplies these data points to other users for entering topology information. Eventually, meta data, such as the type of street, the number of lanes, or the maximum speed for each track can be added to this geodata to arrive at a full-featured map. Aside from this process, the project infrequently imports raw data based on donations from commercial companies and governmental data sets, such as TIGER and Landsat 7.

Map data is further augmented with additional information like house numbers, power lines, amenities, or the shape of buildings. For the latter, height information can be attached using attributes, which has already been done for wide areas. Related to that, there has been work on combining RADAR height information with OpenStreetMap maps to provide 3D scene graphs [19].

In order to clarify the level of granularity at which geodata is made available in OpenStreetMap, and to illustrate how we use this data, the following sections will briefly highlight the XML data format and, by means of an example, the steps we take to make this data available to the microscopic road traffic simulator SUMO.

---

[1] http://www.census.gov/geo/www/tiger/

[2] http://www7.informatik.uni-erlangen.de/veins/

[3] http://www.openstreetmap.org

A. XML data format

Geodata from the OpenStreetMap web page is, by default, exported as XML data. The format of these .osm files has not been formally specified, however, there are three predefined data primitives from which all objects on a map can be built:

- *Nodes* are the basic elements of an OSM map. They describe a single point on a map, defined (at a minimum) by the XML attributes `lat` and `lon` for latitude and longitude. Additionally, each node is assigned a unique `id`.
- A *way* describes a linear feature and can be seen as a graph $G = (V, E)$ with $V = (v_1, v_2, .., v_n)$ being nodes and $2 \leq |V| \leq 2000$. Ways can be closed ($v_1 = v_n$) or open ($v_1 \neq v_n$). A way therefore has to have an ordered list of all connected nodes referred by their respective id. Assignment of road or area type is not done in the single nodes but in the way element itself.
- *Relations*, available since October 2007, describe groups of data primitives and express a relation between them, as well as their role. A straight-forward example would be a relation between ways forming a route. Another application scenario is the modeling of boundaries consisting of closed ways (areas).

All additional information is attached to the data primitives by key value pairs within the XML element `tag`. There are no restrictions on which keys or values can be used. A simple example for the usage of those primitives can be found in listing 1.

Listing 1. Example excerpt from an .osm file

```
<?xml version="1.0" encoding="utf-8"?>
<osm version="0.6" generator="OpenStreetMap server">
  <node id="1" lat="11.5000074" lon="9.1580008"/>
  <node id="2" lat="11.5000059" lon="9.1580053"/>
  <node id="3" lat="11.5000055" lon="9.1580040">
    <tag k="barrier" v="gate"/>
  </node>
  <way id="27">
    <nd ref="1"/>
    <nd ref="2"/>
    <nd ref="3"/>
    <tag k="highway" v="residential"/>
    <tag k="maxspeed" v="50"/>
    <tag k="lanes" v="1"/>
  </way>
</osm>
```

B. Importing OSM into SUMO

When importing OSM data into SUMO for traffic simulation, there are certain steps that have to be taken. First of all, the map projection has to be chosen as SUMO operates on planar maps. A common choice would be the conformal, angle-preserving Mercator projection, or UTM, a zone-based variant. To start the conversion of .osm data to SUMO maps, the SUMO *netconvert* tool can be used: In the presented example, we use the UTM projection based on the WGS 84 data and reference ellipsoid and configure default values for missing attributes, e.g., of primary roads:

Listing 2. Example mymap.edgetypes.xml

```
<types>
  <type id="highway.primary" priority="9" nolanes="2"
      speed="20"/>
</types>
```

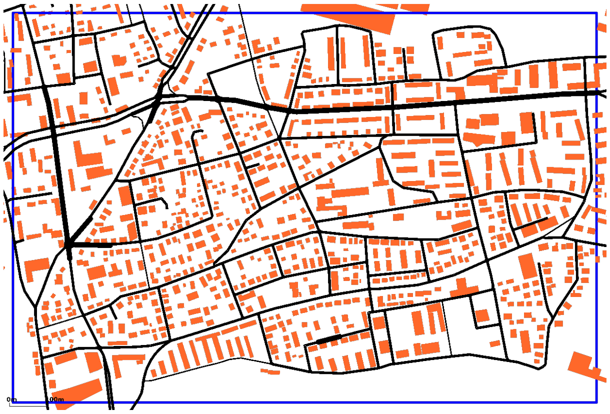


Fig. 2. Screenshot of SUMO showing the considered ROI in the full-featured simulation scenario.

After this conversion, every road segment will be represented in SUMO by one `edge` element per direction, as long as it does not connect to other roads. Junctions in the road network will form nodes which are connected by these edges. Every lane of a road will be represented by a `lane` element with stored attributes, such as its maximum speed:

Listing 3. Excerpt from mymap.net.xml

```
<edge id="-2#0" from="40" to="50" priority="-1" type=""
     function="normal">
  <lanes>
    <lane id="-2#0_0" depart="1" maxspeed="8.33"
         length="184.33" shape="2215.90,3932.27
         2356.69,3813.29"/>
    <lane id="-2#0_1" depart="0" maxspeed="8.33"
         length="184.33" shape="2218.03,3934.79
         2358.82,3815.81"/>
  </lanes>
</edge>
```

SUMO netconvert will only import roads, meaning additional information, such as data on buildings or other areas is discarded. While for some simulation scenarios these are not needed they are of importance for others, especially ones using obstacles for radio dissemination.

Arbitrary polygons can be modeled in SUMO by using the `poly` XML element in one of the files linked in the configuration file. To import this data from .osm files (and, again, perform the necessary projection of coordinates), the SUMO *polyconvert* tool can be used, making them available to SUMO and, via TraCI, to the network simulator. A typemap can be used to specify all types that should be selected for export. Listing 4 shows an example for extracting building information from .osm files. It also contains additional information on how these areas should be visualized in SUMO.

Listing 4. Example mymap.polytpes.xml

```
<polytypes>
  <polytype id="building.yes" name="building"
       color="0.00,1.00,0.00" fill="true" layer="-1"
       discard="false" />
</polytypes>
```

The resulting mymap.poly.xml will then contain all buildings from the .osm file. Listing 5 shows an example that models one building and a polygon representing the Region of Interest (ROI) for the simulation.

Listing 5. Excerpt from mymap.poly.xml

```
<polys>
  <poly id="-150412472" type="building"
       color="1.00,0.00,0.00" fill="1" layer="-1"
       shape="6337.46,4869.64 6332.41,4883.59
       6321.98,4880.30 6327.69,4866.37 6337.46,4869.64"/>
  <poly id="roi" type="roi" color="0,0,1" fill="false"
       layer="-1" shape="6100,4100 6100,5100 7600,5100
       7600,4100 6100,4100"/>
</polys>
```

After the previous steps have been executed, all relevant geodata is available to SUMO; its graphical representation of the imported map would look as depicted in Figure 2.

IV. SIMULATIVE EVALUATION

We rely on our Veins simulation framework to set up and simulate three different scenarios, corresponding to three different degrees of availability of geodata, based on OpenStreetMap data of the city of Ingolstadt. For the first scenario, we import only road geometry and class information, thus relying on the extrapolation features of SUMO to assemble a basic simulation scenario. In the second configuration, we import all of the available information on the road network of Ingolstadt, thus arriving at a scenario with accurate lane counts, speed limits, traffic lights, and turn restrictions. Finally, we further augment the second configuration with OpenStreetMap data on buildings to assemble a third scenario where radio transmissions will be attenuated by these obstacles.

In order to evaluate the degree to which accurate IVC simulations depend on the availability of more than just road topology, we examine two metrics: the *neighbor count*, i.e. the number of nodes in communication range, and the *neighbor lifetime*, i.e. the duration of contact with nodes in communication range.

Traffic flows for all scenarios were computed by randomly generating origin/destination pairs and iteratively applying the Dynamic User Assignment (DUA) algorithm of [20], as implemented in SUMO, until it reported a stable, optimal distribution of flows. While traffic is simulated in the whole city of Ingolstadt, we simulate network communication only for nodes within a smaller Region of Interest (ROI). In the evaluation, we focus on the 1.5 km × 1 km ROI shown in Figure 2, which contains a typical mixture of high- and low-capacity roads, traffic lights, and unregulated intersections.

Each of the three scenarios was simulated for radios of four different power levels, corresponding to maximum transmission ranges of approx. 140 m, 180 m, 240 m, and 1 200 m.

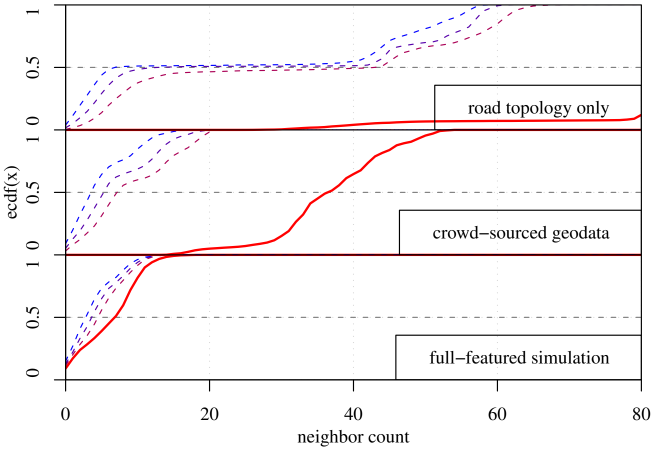


Fig. 3. Distribution of number of nodes in communication range. Shown are results for radios of different power (results for a maximum range of approx. 1.2 km are represented by the red, solid line) and map data of different quality. Lower-quality map data can lead to unrealistically dense car clusters.

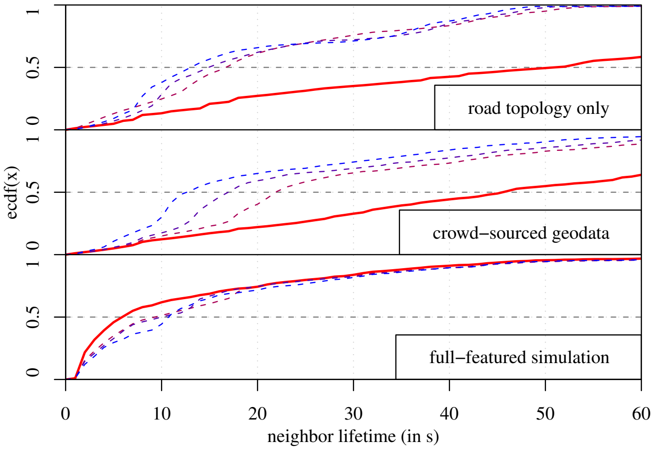


Fig. 4. Distribution of duration of contact with nodes in communication range. Shown are results for radios of different power and map data of different quality. Lower-quality map data can suggest overly long and differently-distributed contact durations.

Figure 3 illustrates the results of these simulations by plotting the neighbor count of nodes in the form of an empirical cumulative density function (eCDF). Thus, we can focus on the distributions of individual measurements. As can be seen from the graphs, the distributions vary widely, depending on the level of detail with which the scenarios had been modeled. Looking at the neighbor count in a road network based on topology data alone, a strongly expressed bimodality in its distribution can be observed: in this network, vehicles alternately drive in freely flowing traffic and in heavily congested traffic; this is directly reflected in the neighbor count. This effect is indicative of the mismatch between traffic demand and road network capacity due to the imperfect extrapolation of lane counts and speed limits performed by SUMO; some roads had an unnecessarily high lane count, while other areas of the network could not cope with the number of vehicles passing through, leading to the observed congestions. Employing a road network based on crowd-sourced geodata, this issue is resolved, which is also directly reflected in a more regular distribution of neighbor counts. The inclusion of data on buildings in a full-featured simulation had only little impact on the distribution of neighbor counts if low-power radios were used. As was to be expected, however, if radios were able to transmit over 1.2 km, the impact of radio obstacles was significant (solid red line in Figure 3). Not considering them in the simulation led to unrealistically high neighbor counts – almost as high as in scenarios based on road topology alone.

Looking at the neighbor lifetime (Figure 4), a core metric of network stability, a reverse trend can be observed – including geodata on road features (but not on buildings) in

IVC simulations increased the simulated networks' stability; yet, when also including geodata on buildings, the network stability turned out to be worse. With increasing transmission power this decrease in network stability becomes increasingly pronounced, as buildings frequently obstruct radio transmissions across roads, thus frequenly chopping what used to be a stable connection between two vehicles into multiple short-lived connections.

V. CONCLUSION AND FUTURE WORK

Based on the presented simulation study, we conclude that including information beyond the mere topology of roads in IVC simulations can have a profound impact on core network metrics, both in terms of aggregate statistics and shape of distribution. Freely available crowd-sourced geodata can be used to fill in these blanks and provide this information. We argue that, this way, more accurate results can be generated by IVC simulations at no cost to fellow researchers. The results of this study also illustrate the importance of basing simulative studies of IVC protocols on realistic scenarios.

This also touches on issues completely different from the aforementioned network and road traffic models: inaccurate demand models will lead to artificial bottlenecks appearing in the road network. In our simulation, we took care to compare values from simulated induction loops with real values provided by the local authorities, but road traffic was still ultimately based directly on the offered road network capacity, not on actual demands. Accurate demand modeling is a task being studied in the transport and traffic engineering community. We thus consider the creation and use of unified high-quality reference scenarios an important next step in IVC research.

REFERENCES

[1] T. L. Willke, P. Tientrakool, and N. F. Maxemchuk, "A Survey of Inter-Vehicle Communication Protocols and Their Applications," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 2, pp. 3–20, 2009.

[2] IEEE 802.11 Working Group of the IEEE 802 Committee, "802.11p/D10.0 - Amendment 7: Wireless Access in Vehicular Environments," Tech. Rep., February 2010.

[3] L. Wischhof, A. Ebner, and H. Rohling, "Information dissemination in self-organizing intervehicle networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 1, pp. 90–101, March 2005.

[4] J. Rybicki, B. Scheuermann, W. Kiess, C. Lochert, P. Fallahi, and M. Mauve, "Challenge: Peers on Wheels - A Road to New Traffic Information Systems," in *13th ACM International Conference on Mobile Computing and Networking (ACM MobiCom 2007)*, Montreal, Canada, September 2007, pp. 215–221.

[5] E. Brockfeld and P. Wagner, "Testing and Benchmarking of Microscopic Traffic Flow Simulation Models," in *10th World conference on transport research (WCTR)*, Istanbul, Turkey, July 2004.

[6] C. Sommer and F. Dressler, "Progressing Towards Realistic Mobility Models in VANET Simulations," *IEEE Communications Magazine*, vol. 46, no. 11, pp. 132–137, November 2008.

[7] K. Ibrahim and M. C. Weigle, "ASH: Application-aware SWANS with highway mobility," in *27th IEEE Conference on Computer Communications (INFOCOM 2008): IEEE Workshop on Mobile Networking for Vehicular Environments (MOVE 2008)*. Phoenix, AZ: IEEE, April 2008.

[8] M. Treiber, A. Hennecke, and D. Helbing, "Congested Traffic States in Empirical Observations and Microscopic Simulations," *Physical Review E*, vol. 62, p. 1805, 2000.

[9] R. Mangharam, D. Weller, R. Rajkumar, P. Mudalige, and F. Bai, "GrooveNet: A Hybrid Simulator for Vehicle-to-Vehicle Networks (Invited Paper)," in *2nd International Workshop on Vehicle-to-Vehicle Communications (V2VCOM 2006)*, San Jose, CA, July 2006.

[10] D. R. Choffnes and F. E. Bustamante, "An Integrated Mobility and Traffic Model for Vehicular Wireless Networks," in *2nd ACM International Workshop on Vehicular Ad hoc Networks (VANET 2005)*. Cologne, Germany: ACM, September 2005, pp. 69–78.

[11] D. Krajzewicz, G. Hertkorn, C. Rössel, and P. Wagner, "SUMO (Simulation of Urban MObility); An open-source traffic simulation," in *4th Middle East Symposium on Simulation and Modelling (MESM 2002)*, Sharjah, United Arab Emirates, September 2002, pp. 183–187.

[12] C. Sommer, Z. Yao, R. German, and F. Dressler, "On the Need for Bidirectional Coupling of Road Traffic Microsimulation and Network Simulation," in *9th ACM International Symposium on Mobile Ad Hoc Networking and Computing (Mobihoc 2008): 1st ACM International Workshop on Mobility Models for Networking Research (MobilityModels 2008)*. Hong Kong, China: ACM, May 2008, pp. 41–48.

[13] M. Piorkowski, M. Raya, A. L. Lugo, M. Grossglauser, and J.-P. Hubaux, "Joint Traffic and Network Simulator for VANETs," in *MICS 2006, Poster Session*, Zurich, Switzerland, October 2006.

[14] C. Sommer, R. German, and F. Dressler, "Bidirectionally Coupled Network and Road Traffic Simulation for Improved IVC Analysis," *IEEE Transactions on Mobile Computing*, 2010, available online: 10.1109/TMC.2010.133.

[15] L. Breslau, D. Estrin, K. Fall, S. Floyd, J. Heidemann, A. Helmy, P. Huang, S. McCanne, K. Varadhan, Y. Xu, and H. Yu, "Advances in network simulation," *IEEE Computer*, vol. 33, no. 5, pp. 59–67, May 2000.

[16] A. Varga, "The OMNeT++ Discrete Event Simulation System," in *European Simulation Multiconference (ESM 2001)*, Prague, Czech Republic, June 2001.

[17] A. Wegener, M. Piorkowski, M. Raya, H. Hellbrück, S. Fischer, and J.-P. Hubaux, "TraCI: An Interface for Coupling Road Traffic and Network Simulators," in *11th Communications and Networking Simulation Symposium (CNS'08)*, Ottawa, Canada, April 2008.

[18] F. Dressler and C. Sommer, "On the Impact of Human Driver Behavior on Intelligent Transportation Systems," in *71st IEEE Vehicular Technology Conference (VTC2010-Spring)*. Taipei, Taiwan: IEEE, May 2010.

[19] M. Over, M. Elhachadi, A. Schilling, and A. Zipf, "Automatische Generierung von Detailstufen für ein länderübergreifendes, standardkonformes 3D Stadt- und Landschaftsmodell am Beispiel von OSM-3D.org," in *Symposium für Angewandte Geoinformatik, AGIT*, Salzburg, Austria, July 2010.

[20] C. Gawron, "Simulation-Based Traffic Assignment – Computing User Equilibria in Large Street Networks," PhD Thesis, University of Cologne, 1999.