



# A large-scale vehicular mobility dataset of the Cologne urban area

Sandesh Uppoor, Marco Fiore

## ► To cite this version:

Sandesh Uppoor, Marco Fiore. A large-scale vehicular mobility dataset of the Cologne urban area. 14èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications (AlgoTel), May 2012, Hérault, France, France. pp.1-4. hal-00690269

**HAL Id: hal-00690269**

**<https://hal.science/hal-00690269>**

Submitted on 23 Apr 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# *A large-scale vehicular mobility dataset of the Cologne urban area<sup>†</sup>*

Sandesh Upoor<sup>1</sup> and Marco Fiore<sup>1,2</sup>

<sup>1</sup>*INRIA*

<sup>2</sup>*INSA-Lyon, CITI, F-69621, Villeurbanne, France*

---

Vehicles are envisioned to become real communication hubs in the near future, thanks to the growing presence of radio interfaces on the cars as well as to the increasing utilization of smartphones and tablets by their passengers. The single most distinguishing feature of vehicular networks lies in the mobility of users, which is the result of the interaction of complex macroscopic and microscopic dynamics. Notwithstanding the improvements that vehicular mobility modeling has undergone during the past few years, no car traffic trace is available today that captures both macroscopic and microscopic behaviors of drivers over a large urban region, and does so with the level of detail required for networking research. In this paper, we present a realistic synthetic dataset of the car traffic over a typical 24 hours in a 400-km<sup>2</sup> region around the city of Köln, in Germany. We outline how our mobility description improves today's existing traces and show the potential impact that a comprehensive representation of vehicular mobility can have on the evaluation of networking technologies.

**Keywords:** Vehicular networks, mobility, connectivity analysis

---

## 1 Introduction

Simulation remains the mean of choice for the evaluation of large-scale deployments of new vehicular networking solutions. Notably, an ever increasing attention is paid to the faithful simulation of the unique dynamics of car mobility, as it is today commonly agreed that the high-speed, strongly-correlated and constrained movements of vehicles can dramatically affect the network performance.

The challenge lies in generating traffic traces that : (i) compass very large urban areas, i.e., whole cities including their surroundings ; (ii) present realistic microscopic mobility features, i.e., that properly reproduce the movement of individual drivers in presence of other cars, traffic lights, road junctions, speed limits, etc. ; (iii) are realistic also from a macroscopic point of view, i.e., that faithfully mimic the evolution of large traffic flows across a metropolitan area over time.

Currently, the vehicular mobility traces that are commonly employed for the validation of network protocols and solutions are limited to a small area and short duration [RB08, CB09] or large-scale and accounting for macroscopic mobility, but lacking microscopic detail, in terms of traffic, time and space granularity [NC03, Ni09].

Later sections in this paper explain the original Köln dataset and its improvements and we prove the realistic nature by comparing with other well known vehicular mobility traces.

## 2 The TAPASCologne dataset

The vehicular mobility dataset we introduce in this short paper is mainly based on data made available by the TAPASCologne project [TAP], an initiative of the Institute of Transportation Systems at the German Aerospace Center (ITS-DLR). In order to generate the trace, different state-of-art data sources and simulation tools are brought together as discussed below.

---

<sup>†</sup>This work was supported by the joint lab between INRIA and Alcatel-Lucent Bell Labs on Self Organizing Networks..

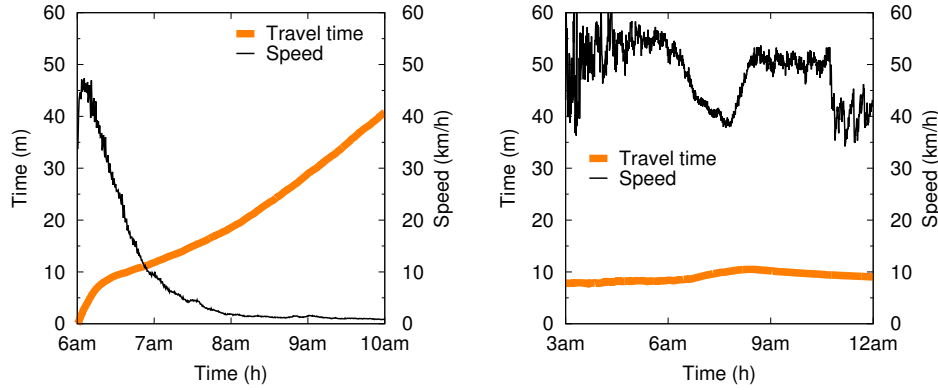


FIGURE 1: Traffic features of the TAPASCologne dataset before (top) and after repair (bottom)

The street layout of the Köln urban area is obtained from the OpenStreetMap (OSM) database [OSM]. The OSM project provides freely exportable maps of cities worldwide, which are contributed and updated by a vast user community. The OSM road information is generated and validated by means of satellite imagery and GPS traces, and is commonly regarded as the highest-quality road data publicly available today.

The microscopic mobility of vehicles is simulated with the Simulation of Urban Mobility (SUMO) software [SUM], an open-source, space-continuous, discrete-time traffic simulator capable of accurately modeling the behavior of individual drivers, accounting for car-to-car and car-to-roadsign interactions. The microscopic models implemented by SUMO are Krauss' car-following model and Krajzewicz's lane-changing model. These models have been long validated by the transportation research community, a fact that, jointly with the high scalability of the simulator, makes of SUMO the most complete and reliable among today's open-source microscopic vehicular mobility generators.

The macroscopic traffic flows across the Köln urban area are derived through the Travel and Activity Patterns Simulation (TAPAS) methodology. This technique generates the trips of each driver by exploiting information on (i) the population, i.e., home locations and socio-demographic characteristics, (ii) the points of interests in the urban area, i.e., places where working and free-time activities take place, and (iii) the time use patterns, i.e., habits of the local residents in organizing their daily schedule [GH04]. Within the context of the TAPASCologne project, the TAPAS methodology was applied on real-world data collected in the Köln region by the German Federal Statistical Office, including 30,700 daily activity reports from more than 7000 households. The resulting traffic flows faithfully mimics the daily movements of inhabitants of the area for a period of 24 hours, for a total of 1.2 million individual trips.

The individual components presented above are combined in order to generate the vehicular mobility dataset. Unfortunately, the result obtained by directly running the vehicular mobility simulation with the data sources made available by OSM and TAPASCologne leads to a plain unusable result, as proven by the top plots of Fig. 1. The left image shows the average speed of vehicles in the simulation degrading to zero which is quite unrealistic in a city of Köln thereby affecting the travel time indicating huge traffic jam, where most of the drivers are stuck and cannot complete their trip, or even enter the simulation area.

In order to make the dataset usable, we had to repair it, identifying and solving a number of issues, that are detailed next.

**Macroscopic traffic demand.** The demand in the TAPASCologne dataset is not limited to the vehicular traffic; rather, it includes information on the daily trips of all Köln inhabitants, independently from whether their walk to their destination, or employ public transports, or take a car as either passengers or drivers. Clearly, we are only interested in the latter kind of mobility, since the volume of vehicular traffic directly maps to that of car drivers. According to [GH04, Fig. 4], car drivers account for approximately 50% of the overall trips in the TAPASCologne O/D matrix: thus, we adjusted the O/D matrix by only considering that one trip every two concerns the movement of a vehicle. Additionally, the original demand is smoothened to mitigate the unrealistic variability in the injected traffic over short time scales, which was verified to be a

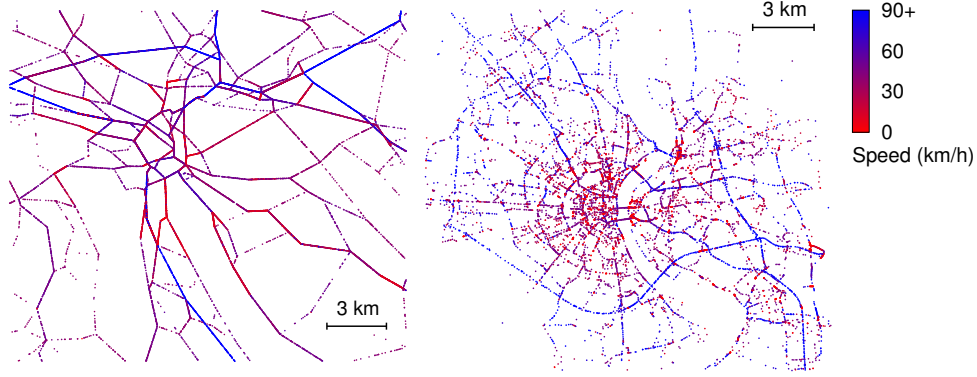


FIGURE 2: Trace snapshot (best viewed in colors)

major cause of congestion.

**OpenStreetMap data.** A second source of errors in the simulation was identified in the OSM road topology. Although very complete, the information embedded in the map proved to be at times inconsistent with respect to reality. The impact of such inconsistencies, albeit negligible on most of the usages of OSM, revealed to be dramatic for the simulation of vehicular mobility. The problems we identified were mostly related to wrong traffic movement restrictions being enforced on some road segments. We corrected the OSM data by visually checking the restrictions against the real-world road signs via the Google Street View service.

**Road topology conversion.** The OSM road information is imported by SUMO through an automated conversion process that proved not to be error-free. First, the topological representation of OSM was, at times, simply unfit to be directly converted to the SUMO street layout. That was, e.g., the case of bidirectional roads appearing as two parallel unidirectional roads in OSM : when joining intersections, such roads would create two adjacent junctions instead of the one present in the reality, also duplicating the road signs. Second, a number of roads, containing OSM attributes not recognized by the conversion tool, were just removed. Third, the conversion tool automatically deployed additional traffic lights that had indeed a negative impact on the traffic flow. We thus repaired the OSM data for conversion.

**Traffic assignment.** The traffic assignment defines the way drivers choose the route to reach their intended destination. The basic solution adopted by SUMO, a Dijkstra’s algorithm with road weights determined by the length and maximum speed, resulted in heavy congestion on the fastest roadways. We thus resorted to the traffic assignment technique proposed by Gawron [Gaw98], that, by iteratively moving part of the traffic to alternate, less congested paths, can achieve a so-called dynamic traffic equilibrium.

The result of the repaired dataset is shown in the left plot of Fig. 1. There, the simulated traffic now mimics the normal daily road activity with agreeable speed and travel times, since the fixed road topology can correctly accommodate the updated traffic demand and assignment.

Interestingly, we also found the synthetic traffic to nicely match that observed in the real world, through real-time traffic information services.

### 3 Connectivity analysis

We hint the potential impact of the mobility trace on the performance evaluation of networking protocols, by analyzing its connectivity properties. We compare the TAPASCologne dataset with a trace generated via the Multi-agent Microscopic Traffic Simulator (MMTS) [NC03]. The latter describes car traffic around Zurich, Switzerland : the simulated area, duration and number of trips are close to those of our dataset. That of Zurich is the only large-scale vehicular trace currently available, and is widely employed in the vehicular network literature.

Fig. 2 portrays snapshots of the traffic at 7 :00 am, in the Zurich (left) and Köln (right) scenarios. The level of detail provided by our dataset is clearly higher. Indeed, the Zurich trace is characterized by a coarser road map, only accounting for major traffic arteries. Also, the MMTS is based on a queuing approach, less

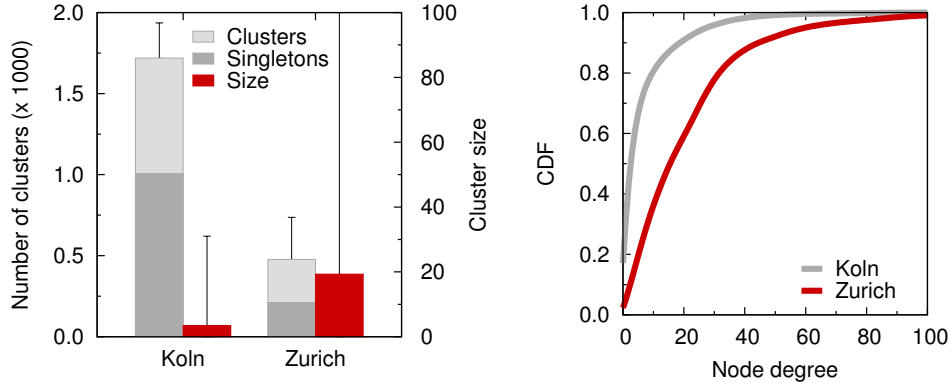


FIGURE 3: Clustering and degree properties

accurate than SUMO car-following model. As a result, vehicle positions (shown as dots colored depending on the current car speed) are more fine-grained in the Köln case, outlining a considerably more precise street layout.

Such a higher microscopic-level accuracy leads to significantly different vehicular network properties, shown in Fig. 3 for a 100-meter transmission range. In the left plot, we can observe that the Zurich trace results in a much more connected network, with vehicles grouping in a lower number of larger clusters, i.e., disconnected components. More precisely, the Zurich trace presents either giant connected components or very small clusters, as indicated by the dramatically high standard deviation of the cluster size. Such giant components cannot instead be found in the Köln scenario, where clusters tend to be smaller and more uniform in size. The reduced connectivity of the Köln trace is confirmed by the right plot, showing that 1-hop neighborhoods are significantly smaller than in the Zurich case. Indeed, 60% of the vehicles have less than five other cars within communication range in the Köln dataset, whereas this percentage drops to 15% in the Zurich trace.

These results let us speculate that a comprehensive description of the vehicular mobility, accounting for high realism at a microscopic and macroscopic level, leads to very different topological properties of the network. In particular, tests conducted on mobility traces characterized by simplistic macroscopic or microscopic modeling appear to result in more connected and stable networks, which could in turn lead to over-optimistic conclusions on the real-world performance of network protocols.

## Références

- [CB09] G. Malnati C. Barberis. Epidemic information diffusion in realistic vehicular network mobility scenarios. In *IEEE ICUMT*, 2009.
- [Gaw98] C. Gawron. An iterative algorithm to determine the dynamic user equilibrium in a traffic simulation model. In *International Journal of Modern Physics C*, pages 393–407, 1998.
- [GH04] P. Wagner G. Hertkorn. The application of microscopic activity based travel demand modelling in large scale simulations. In *World Conference on Transport Research*, 2004.
- [NC03] K. Nagel N. Cetin, A. Burri. A large-scale multi-agent traffic microsimulation based on queue model. In *STRC*, 2003.
- [Ni09] H. Zhu M. Li Y. Zhu L.M. Ni. Hero online real-time vehicle tracking. In *IEEE Transactions on Parallel and Distributed Systems*, pages 740–752, 2009.
- [OSM] OSM. Openstreetmap. <http://www.openstreetmap.org>.
- [RB08] P. Sommer R. Baumann, F. Legendre. Generic mobility simulation framework (gmsf). In *ACM MobilityModels*, 2008.
- [SUM] SUMO. Simulation of urban mobility. <http://sumo.sourceforge.net>.
- [TAP] TAPASCologne. Tapascologne project. <http://sourceforge.net/apps/mediawiki/sumo/index.php?title=TAPASCologne>.