Large-scale Urban Vehicular Mobility for Networking Research

Sandesh Uppoor, Marco Fiore
Université de Lyon, INRIA
INSA-Lyon, CITI, F-69621, Villeurbanne, France
{sandesh.uppoor,marco.fiore}@insa-lyon.fr

Abstract—Simulation is the tool of choice for the large-scale performance evaluation of upcoming telecommunication networking paradigms that involve users aboard vehicles, such as next-generation cellular networks for vehicular access, pure vehicular ad hoc networks, and opportunistic disruption-tolerant networks. The single most distinguishing feature of vehicular networks simulation 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^2 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 one the evaluation of networking technologies.

I. INTRODUCTION

Vehicular environments have become increasingly attractive to the telecommunication networking research community over the last years. The reason is that cars are envisioned to become real communication hubs in the near future, thanks to the proliferation of smartphones and tablets, whose Internet-connection capabilities appear especially appealing to passengers aboard cars, as well as to the growing presence of radio interfaces on the vehicles themselves.

Enhanced infrastructure-based systems, involving, e.g., the WiMAX and LTE-A technologies, and novel communication paradigms, such as, e.g., ad hoc and opportunistic networking, are being studied in order to accommodate the traffic generated and requested by forthcoming communicating vehicles. Most of these solution require large-scale performance evaluations that are not feasible through experimentation directly, due to costs and complexity. Simulation becomes thus the tool of choice to assess the quality such solutions.

When simulating a vehicular network, particular attention must be paid to faithfully represent the unique dynamics of car mobility, characterized at a time by high-peak high-variance speeds, road topology-and road rules-constrained movements, and strong movement correlations over time and space. These properties are the result of macroscopic and microscopic car traffic dynamics, that need to be properly modeled in order to perform a simulative campaign whose results are credible.

The relevance of mobility modeling in the simulation of vehicular networks is widely acknowledged, a factor that has led to a substantial progress in the quality of car movement traces for vehicular networking research. The simplistic stochastic models employed in early works [1], [2] have been replaced by random mobility over *realistic road topologies* [3] at first, and by *microscopic vehicular models* borrowed from transportation research [4] later on. These features were then included in dedicated simulation environments, and integrated with road signalization [5], [6]. Since then, vehicular mobility simulators have been growing their complexity and features [7], allowing to accurately simulate the individual movement of vehicles over realistic road topologies.

Today's challenge lies in generating traffic traces that (i) compass very large urban areas, i.e., whole cities including their surroundings, and (ii) are realistic also from a macroscopic point of view, i.e., that faithfully mimic large traffic flows across a metropolitan area. To that end, one must correctly identify the *traffic demand*, i.e., the start time, the origin and the destination of each car trip in the simulated region, which are stored in a so-called Origin/Destination (O/D) matrix. Then, an appropriate *traffic assignment* model needs to be run on the O/D matrix, so to identify the realistic route followed by each driver to reach his/her destination. Indeed, the current common practice is to employ a random traffic demand and shortest path-based assignment, which can lead to unrealistic flows and thus biased simulation results.

The few large-scale vehicular mobility traces accounting for macroscopic mobility and publicly available have been extensively used in the literature, but are either representative of a subset of the whole traffic, as in the case of real-world taxi traces [8], or only model major arterial roads and neglect most of the streets present in urban areas [9]. Additionally, both these kinds of dataset lack detail, since they report car positions with low granularity, in the order of tens of seconds, or employ fast but approximate microscopic mobility models. Recently, the iTetris initiative [10] generated very detailed urban mobility traces based on realistic macroscopic data collected in the city of Bologna, Italy. However, such traces only cover a short timespan of one hour and are limited to relatively small areas of around 10 km^2; moreover, the traces are not publicly available at the moment.

In this paper, we combine the four key aspects mentioned above, i.e., a real-world road topology, an accurate microscopic mobility modeling, a realistic traffic demand, and a state-of-art traffic assignment technique, and generate a large-scale synthetic trace of the car traffic in a the city of Köln, Germany. The dataset covers a region of 400 km^2 for a period of 24 hours. We detail the generation process, and then outline how our mobility dataset compares with some reference traces commonly employed in the recent literature. Our analysis suggest that employing incomplete representations of vehicular mobility in the evaluation of networking protocols may indeed result in over-optimistic performance.

II. THE TAPASCOLOGNE DATASET

The vehicular mobility dataset we introduce in this paper is mainly based on the data made available by the TAPAS-Cologne project [11]. TAPASCologne, an initiative by the Institute of Transportation Systems at the German Aerospace Center (ITS-DLR), aims at reproducing, with the highest level of realism possible, car traffic in the greater urban area of the city of Köln, in Germany.

To that end, different state-of-art data sources and simulation tools are brought together, so to cover all of the specific aspects required for a proper characterization of vehicular traffic. In the remainder of this section, we detail the features of the different components, as well as the process through which they are combined to generate the mobility dataset.

A. Road topology

The street layout of the Köln urban area is obtained from the OpenStreetMap (OSM) database [12]. The OSM project provides freely exportable maps of cities worldwide, which are contributed and updated by a vast user community. Maps include information on roads, railways, buildings, and Points of Interests (PoI) such as parks, commercial centers, leisure centers and commercial activities.

In particular, 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. Indeed, the accuracy of OSM street layouts, comprising highways, major urban arteries and minor roads, often matches that of proprietary ones such as, e.g., Google Maps or Mappy, especially for large cities.

We employed the osmosis tool [13] to filter the OSM data and extract the road topology information for an area of approximately 400 km^2 around the urban agglomeration of Köln, thus including almost 4500 km of roads. We then resorted to the Java OpenStreetMap Editor (JOSM) [14] to repair the OSM data file and make it compatible with the microscopic mobility simulator, as detailed in Sec. III.

B. Microscopic vehicular mobility

The microscopic mobility of vehicles is simulated with the Simulation of Urban Mobility (SUMO) software [15]. SUMO is an open-source, space-continuous, discrete-time traffic simulator developed by the German Aerospace Center (DLR), capable of accurately modeling the behavior of individual drivers, accounting for car-to-car and car-to-road signalization interactions. More precisely, SUMO can import road maps in multiple formats, including OSM, and faithfully reproduce traffic lights, roundabouts, stop and yield signs. The microscopic models implemented by SUMO are Krauss' car-following model [16] and Krajzewicz's lane-changing model [17], that respectively regulate a driver's acceleration and overtaking decisions, by taking into account a number of factors, including the distance to the leading vehicle, the traveling speed, and the acceleration and deceleration profiles. 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 version we employed for the dataset generation is 12.3.
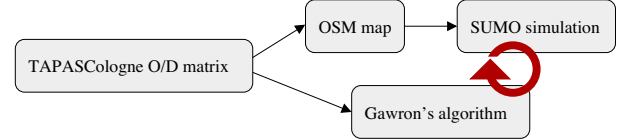

Fig. 1. Simulation workflow

C. Traffic demand

The traffic demand information on the macroscopic traffic flows across the Köln urban area that we employ are derived through the Travel and Activity PAtterns Simulation (TAPAS) methodology [18]. This technique generates the O/D matrix 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 [19]. Within the context of the TAPASCologne project, the aforementioned 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 [20], [21]. The resulting O/D matrix 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 TAPASCologne O/D matrix is, to the best of our knowledge, the only realistic traffic demand dataset of a large urban region available to date.

D. Traffic assignment

The actual assignment of the vehicular traffic flows described by the TAPASCologne O/D matrix over the road topology is performed by means of Gawron's algorithm [22]. This traffic assignment technique computes the fastest route for each vehicle, and then assigns to each road segment a cost reflecting the intensity of traffic over it. By iteratively moving part of the traffic to alternate, less congested paths, and recomputing the road costs, the scheme finally achieves a so-called user equilibrium. Additionally, since the intensity of the traffic demand varies over a day, the traffic assignment model must also be able to adapt to the time-varying traffic conditions. Indeed, Gawron's algorithm satisfies such a requirement, thus attaining a so-called dynamic user equilibrium. Gawron's is one the most popular traffic assignment techniques developed within the transportation research community, and allows to reach a road capacity utilization close to reality and significantly higher than that obtained with, e.g., a simple weighted Dijkstra algorithm.

E. Simulation

The individual components presented above must be combined in order to generate the vehicular mobility dataset. The simulation workflow is depicted in Fig. 1. First, the information contained in the TAPASCologne O/D matrix are used to identify the boundaries of the exact simulation region, extract the associated map from OSM and filter it so to remove unneeded content that does not concern the road layout. Then the OSM map is converted to a format readable by SUMO, and fed to the microscopic mobility simulator.
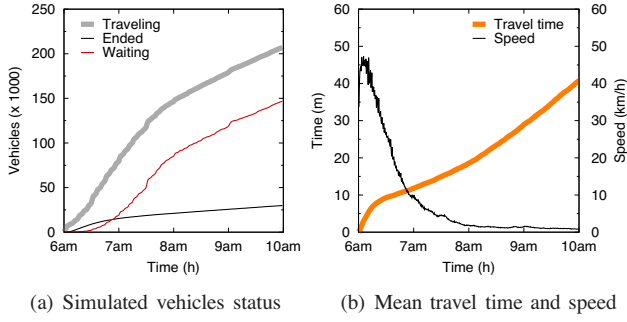
(a) Simulated vehicles status   (b) Mean travel time and speed

Fig. 2.   Original TAPASCologne dataset. Traffic features over time

(a) Original   (b) Fixed

Fig. 4.   Volume of traffic injected in the road network between 6:00 am and 12:00 am, according to the TAPASCologne O/D matrix
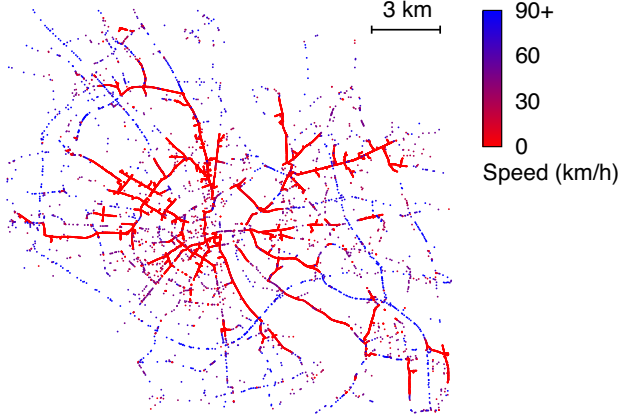

Fig. 3.   Original TAPASCologne dataset. Snapshot of the traffic status at 7:00 am, in a 400 km^2 region centered on the city of Köln. Blue vehicles are moving, whereas bright red ones are still (this figure is best viewed in colors)

The TAPASCologne O/D matrix is also used as an input to Gawron's algorithm, which, in turn, determines an initial traffic assignment and provides it to SUMO. Then, a first vehicular mobility simulation can be run with SUMO, and, once finished, a feedback on the resulting traffic density over the road topology is sent back to Gawron's algorithm. Based on such new information, a new traffic assignment is computed, and a second SUMO simulation is run. The process is repeated until a traffic assignment is generated that allows to sustain the whole volume of the traffic demand.

III. REPAIRING THE DATASET

The result obtained by running the vehicular mobility simulation with the data sources made available by OSM and TAPASCologne is plain unusable. In Fig. 2(a), we plot the evolution over time of the number of vehicles that (i) are traveling on the road topology, (ii) have successfully ended their trip by reaching the their destination, (iii) are waiting to enter the road topology, which they cannot presently do due to an excessive congestion of the road segment they are supposed to start their trip from. This third condition is a simulation artifact, identifying situations where the road topology cannot accommodate all the traffic demand in the O/D matrix, and, as such, is an undesirable effect. Clearly, the three sets of vehicles above are disjoint.

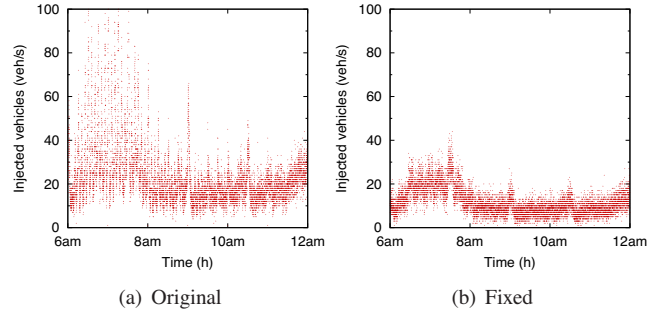From the plot, we can note how the number of traveling vehicles present in the simulation rapidly grows up to exceed

a hundred thousands units, a figure completely unrealistic for a city the size of Köln. Additionally, such a number does not tend to decrease as one could expect once the morning traffic peak is exhausted; instead, it keeps growing indefinitely. It is also possible to observe that the number of vehicles that end their trip grows very slowly over time: in fact, from the values portrayed in the figure, only a very small fraction of the cars that are present on the road topology can reach their destination. Finally, the number of vehicles that are waiting to enter the road topology, which we would like to stay as close as possible to zero, consistently grows over time instead.

The mean travel time, in Fig. 2(b), also shows a quite unrealistic behavior, as more than half an hour is required, on average, for a driver to reach its destination at 10:00 am, when the traffic should be sparse. Similarly, the average speed of vehicles, in the same plot, tends to zero as the time elapses.

These results are clear symptoms of how the road topology cannot sustain the volume of cars injected according to the traffic demand model. Indeed, when looking at a snapshot of the car traffic in the region, it is clear that the simulation quickly reduces to a huge traffic jam. As an example, Fig. 3 depicts the map of the road topology at 7:00 am, a daytime typically characterized by rather fluid traffic conditions. However, the road topology is mostly covered by bright red dots, representing cars stuck in heavily congested traffic.

In the following we discuss the reasons for such a simulation result, and present solutions to them.

A. Over-comprehensive and bursty traffic demand

The original TAPASCologne O/D matrix results in the traffic demand volume depicted in Fig. 4(a), which shows the number of vehicles injected in the whole road network every second. By analyzing the O/D matrix, its source data, and its effect on the microscopic mobility simulation, we identified and fixed the following two problems.

First, the demand in the O/D matrix 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 [19, 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.
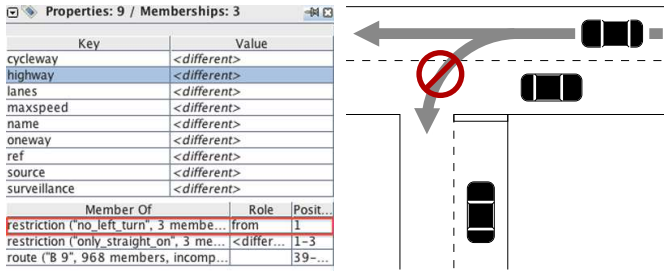
Fig. 5.   Example of wrong restriction enforcement in OpenStreetMap data

Fig. 6.   Example of continuous restriction enforcement in OpenStreetMap

Second, the original demand presented an unrealistic variability in the injected traffic over short time scales. This can be observed in Fig. 4(a), where, within the span of a few minutes, peaks up to 200 vehicles/s in the injected traffic alternate with instants of reduced injected traffic as low as 10 vehicles/s. Such an excessive burstiness is hardly observable in the reality, especially considering that the injected traffic is aggregated over a very large area. Indeed, we observed that these peaks of traffic were a major cause of congestion, forcing large masses of cars to try to enter the road topology at a time, and thus creating traffic jams out of nowhere. In order to address this issue, we smoothed down the original O/D matrix, by adding to the departure time of each vehicle a random offset uniformly distributed in the interval $[-5, 5]$ minutes. This allowed to remove the injection bursts, yet retaining the traffic demand properties over larger time scales. The latter effect can be observed in Fig. 4(b), depicting the injected traffic volume after scaling and smoothing: the overall trend of the plot is the same, clearly resulting in higher traffic between 7:00 am and 8:00 am, and the significant macroscopic peaks of traffic are still present, e.g., at 7:30 am or 9:00 am.

B. Inconsistent road information

A second source of errors in the simulation was identified in the OpenStreetMap 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 OpenStreetMap, revealed to be dramatic for the simulation of vehicular mobility.

A first type of inconsistency is represented by wrong traffic movement restrictions enforced on some road segment. Consider the situation in Fig. 5: there, a *no left turn* restriction (left) is applied, in the OpenStreetMap road information, to the east-west lane of the horizontal road (right). This prevents a car traveling along such a lane to turn left, as in the example in the figure. The OpenStreetMap data contains at times restrictions of this kind which are not actually there in the real world: the shape of the street layout is not affected by such errors, however the microscopic traffic simulation is, since they cause vehicles to get stuck by waiting indefinitely for a possibility to turn and continue their travel. We identified wrong restrictions by checking the features of congested intersections and T-junctions against the real-world road signalization through the Google Street View service: when necessary, we corrected the OpenStreetMap data according to the visual inspection. This allowed us to fix approximately one thousand erroneous restrictions over the road topology in the area under study.
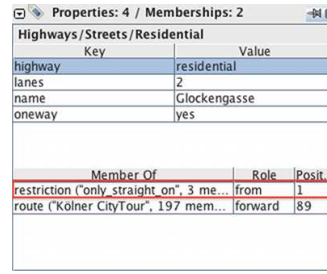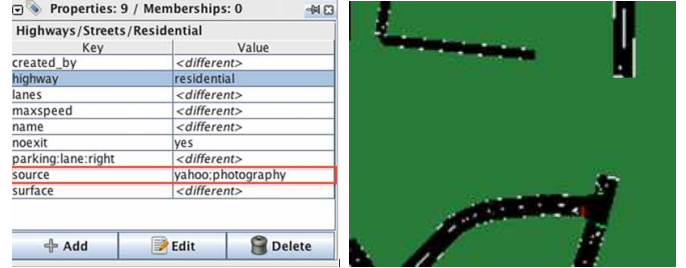

Fig. 7.   Example of unrecognized information ignored during map conversion

We then identified a second type of inconsistency, represented by correct movement restrictions being enforced on some roads, but not applying to the whole road all the time. Fig. 6 portrays an example of such a situation, where two one-way roads, going from west to east and from north to south, respectively, cross each other. In the real world, the roads pass one over the other, and vehicles traveling on the horizontal road can join the southbound traffic flow by means of the slanting relief route. In the OpenStreetMap road representation, the horizontal road is formed by a sequence of segments joined by links (respectively depicted as grey thick lines and red crosses on the right plot); links allow to represent crossings with other roads in the area. More precisely, the horizontal road is tagged as *only straight on* (left), a restriction that affects all of its segments: this correctly forces cars to proceed straight at the bridged intersection with the vertical road. However, the same restriction also applies to the previous segments, preventing vehicles from taking the relief route; as a result, the eastbound traffic cannot join the southbound one. Incorrect restrictions of this kind force vehicles to long detours in order to reach their destinations, resulting in a higher traffic volume over the road topology. We identified such situations in most of the interchange nodes among high-speed roads (arterial roads in the city, the freeway ring around downtown Köln, and highways passing close to the urban agglomeration), preventing traffic from correctly switching among such major ways. We solved the problem by separating the segments of a same road and assigning correct restrictions to each of them, repeating the process for approximately 800 roads.

C. Flawed road topology conversion

The OpenStreetMap road information is imported by SUMO through an automated conversion process that proved not to be error-free. A first cause of problems was the presence of road information not recognized by the conversion tool. E.g., attributes with two values were considered as incorrect by
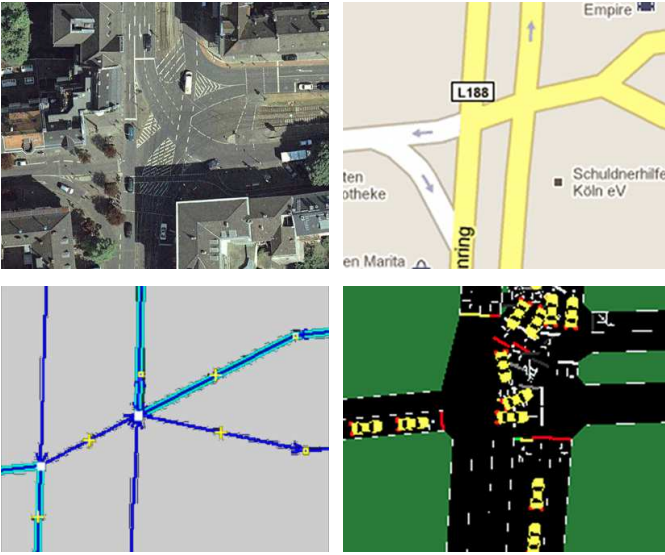
Fig. 8. Example of topological information unfitness during map conversion

(a) Traveling vehicles

(b) Ended trips

Fig. 9. Traffic evolution over multiple iterations of the assignment algorithm

the converter, and the associated roads were not included in the topology used for the simulation. An example is shown in Fig. 7: the double value of the *source* field (left) causes SUMO not to account for the associated road in simulation (right, a road should connect the north and south branches). We corrected the OpenStreetMap data so to make all attributes compatible with the SUMO converter.

A second critical aspect we had to address was the fact that the topological information in OpenStreetMap is, at times, simply unfit to be directly converted to the SUMO street layout. An example is depicted in Fig. 8. There, the real world aerial photography of one intersection (top right), the associated Google map information (top left), and the OpenStreetMap road topology (bottom left) match. However, the conversion of the latter in SUMO results in an exceedingly complex intersection, where vehicles get stuck and rapidly form a permanent traffic jam (bottom right). The reason for such a simulation result is that, since two segment links (white dots in the bottom left plot) are present, SUMO interprets the OpenStreetMap topology as if two road junctions co-existed, one placed right after the other. As a consequence, the number of traffic lights that regulate the car flows to the crossroad is doubled, and yield signs are placed right at the middle of the intersection: the result is the impossibility for vehicular in-flows to correctly merge at the intersection. In order to fix such a problem, we acted directly on the OpenStreetMap road information, by joining road segment links that refer to the same physical intersection. Such an operation allowed then a correct conversion by SUMO, so that no traffic jams were observed anymore at the road junctions. Additionally, we corrected in several cases the number of road lanes entering and leaving an intersection, so to match the aerial photography data.

The third problem we remarked in the OpenStreetMap-to-Sumo conversion lies in the traffic light deployment. Open-StreetMap road information already includes data on the presence or absence of traffic lights at road junctions, and SUMO automatically sets the periodicity of the green and red
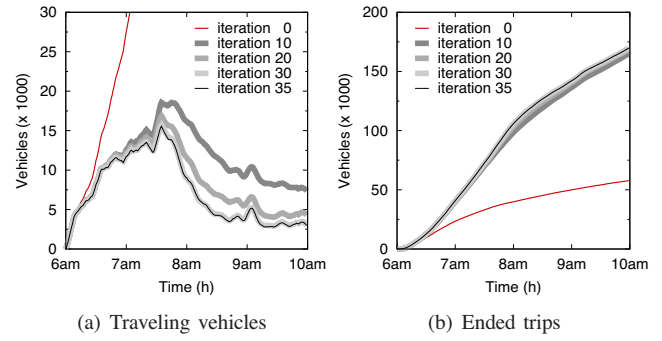
according to the priority of the roads entering each junction. However, the SUMO converter also employs by default a technique to place additional traffic lights over the street layout. After having verified the negative impact of such a traffic light guessing, we disabled it. In addition, we identified a number of situations where the presence of traffic lights was not beneficial, and indeed did not correspond to reality: in particular, this was often the case for intersections formed by peripheral roads with identical priority but very unbalanced traffic. Indeed, the similar periodicity assigned to each road in such a context led to long queues on the trafficked roads. We thus removed such traffic lights from the OpenStreetMap data in order to be consistent with the real-world observations.

D. Simplistic default traffic assignment

Running the microscopic mobility simulation with the traffic demanded corrected as from Sec. III-A and the road topology fixed as from Sec. III-B and Sec. III-C still results in large congestion and continuous traffic jams all over the street layout. The reason lies in the traffic assignment, i.e., the way drivers choose the route to reach their intended destination. Indeed, SUMO employs a simple Dijkstra's algorithm on the road topology graph, by weighting edges, i.e., road segments, on their length, as well as on the maximum speed they allow: clearly, shorter faster roads are preferable, and thus are associated with smaller weights. Unfortunately, this means that drivers having similar origin and destination points will all choose the same routes for their trips: as a result, they will concentrate on major roadways, which will be rapidly filled to their maximum capacity, whereas slower or minor roads will remain unused. Obviously, high-speed roads alone cannot handle the whole demand in the region, and thus the traffic assignment needs to be improved.

To that end, we resorted to the traffic assignment technique proposed by Gawron and presented in Sec. II-D. Such a technique needs to iterate over multiple simulations in order to achieve a dynamic user equilibrium, however the number of iterations cannot be known a priori. Thus, we run the traffic assignment until no significant difference could be observed between subsequent iterations.

Fig. 9 shows the evolution of traffic while iterating the assignment algorithm. The number of vehicles traveling at the same time over the road topology, in Fig. 9(a), tends to explode during the first iterations, as it happened before patching the demand and road topology. However, as Gawron's

(a) Simulated vehicles status   (b) Mean travel time and speed
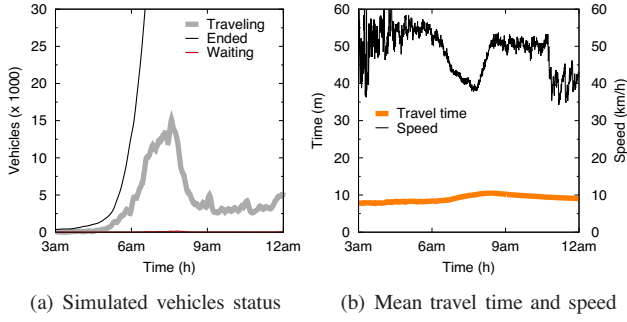
Fig. 10. Final TAPASCologne dataset. Traffic features over time

algorithm iterates, the car traffic is progressively reduced, since drivers tend to employ the different available routes and thus better exploit the capacity of the road network. Fig. 9(b) confirms that iterations significantly improve the traffic conditions, as they increase the number of vehicles that reach their destination, successfully ending their trip. Similar trends were observed for the other traffic metrics, and in all cases iterations after the 35th did not produce any noticeable improvement. Therefore, in the following, we will consider the traffic assignment obtained at the 35th iterations.

IV. LARGE-SCALE URBAN MOBILITY TRACE

The resulting dataset comprises seven hundred thousand trips of vehicles in the Köln larger metropolitan area, over a period of 24 hours. The simulated traffic now mimics the normal daily road activity in the region, as the fixed road topology can accommodate the updated traffic demand and assignment.

Evidences of the correct behavior of the simulated mobility are given in Fig. 10(a). By comparing it to the equivalent plot before repair, in Fig. 2(a), it is clear that the number of traveling cars now follows the traffic demand, being thus very low at night, growing during the peak morning hours, between 7:00 am and 9:00 am, and then remaining steady at a lower values for the rest of the morning. We can remark that the dataset includes approximately 15000 vehicles concurrently traveling over the road topology around 8:00 am. Also, the number of ended trips now grows over time, as more and more drivers reach their destinations over time, and the number of vehicles waiting to enter the simulation is reduced to values close to zero. The average travel time and speed recorded during the morning, in Fig. 10(b) confirm the previous results, as we observe quite constant behaviors, only modified during the peak hours.

As a result, the road traffic at 7:00 am, in Fig. 11, looks significantly better than the original one, in Fig. 3. Indeed, large portions of the roads are in blue, especially in the suburbs, indicating fluid traffic and high traveling speeds. The traffic appears more congested in the city center, as one would expect: however, red areas are dark, and no bright red is visible, meaning that vehicles move at slower speeds, from 30 to 50 km/h, but are not stuck as it was previously the case.

Interestingly, we found the macroscopic traffic simulated in the final TAPASCologne dataset to nicely match that observed in the real world, through real-time traffic information services. In Fig. 12, we compare the traffic information retrieved on
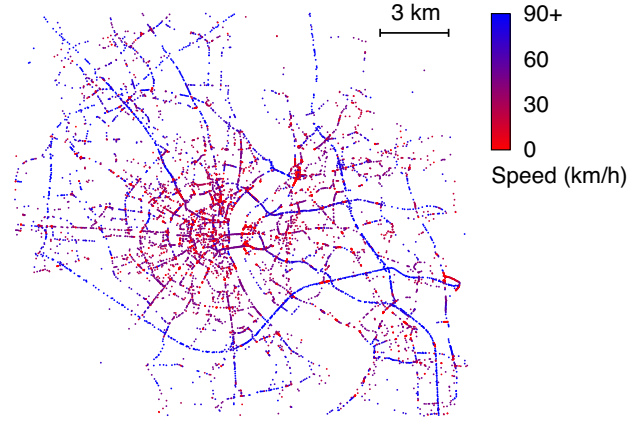

Fig. 11. Final TAPASCologne dataset. Snapshot of the traffic status at 7:00 am, in a 400 km^2 region centered on the city of Köln. Blue vehicles are moving, whereas bright red ones are still (this figure is best viewed in colors)
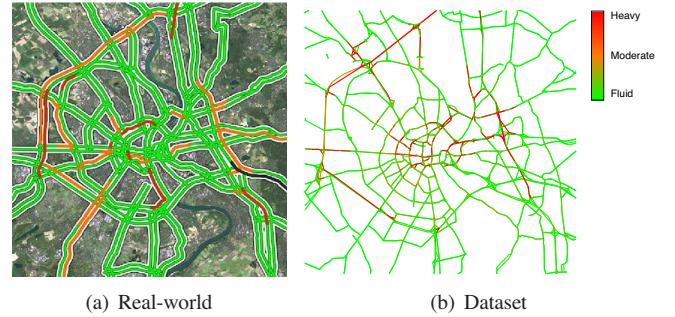

(a) Real-world   (b) Dataset

Fig. 12. Final TAPASCologne dataset. Traffic at 5:00 pm, versus real world

ViaMichelin live traffic website with the simulation output, at 5:00 pm. This represent a critical period of the day, in the middle of the afternoon traffic peak, and key features of real-world mobility patterns are faithfully reproduced in the dataset: e.g., the congestion on the highways around the city, where commuters merge with long-distance travelers passing through the region, or the heavy traffic on the bridges that connect the two parts of the city. Although we acknowledge that more rigorous tests are needed to fully prove the realism of the dataset, we regard the result as a very encouraging start, especially considering that more complex assessment are unfeasible at this moment due to the unavailability of sensible data (e.g., traffic counter records in the urban area).

V. VEHICULAR NETWORK CONNECTIVITY ANALYSIS

In order to provide a preliminary study of the potential impact that a large-scale realistic vehicular mobility trace could have on the performance evaluation of networking protocols, we analyze the connectivity properties of the TAPASCologne dataset, and compare them with those of other vehicular traces employed in the literature. The rationale for this choice is that such an approach is protocol-independent and thus allows us to draw results of general validity. In the following, we will employ a 100 m unit disc model in order to determine the vehicular network connectivity. We reckon that this is a very simplistic approach, however it allows us to evaluate the impact of mobility, which is the focus of this paper, without
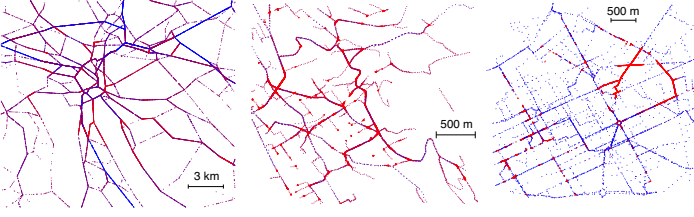
Fig. 13. Road topology and instantaneous speed in the reference scenarios: Zurich region (left), Zurich downtown (middle) and Turin downtown (right)
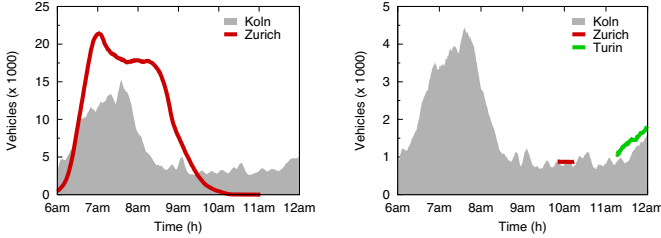

Fig. 14. Number of vehicles over time concurrently traveling in the larger *region* scenarios (left) and smaller *downtown* scenarios (right)

Fig. 15. Average number of clusters, with singletons, and mean cluster size

biases due to the highly irregular signal propagation of urban environments. Including a realistic signal propagation is part of our future objectives, as discussed in Sec. VI.

Reference scenarios. We consider three reference scenarios, portrayed in Fig. 13. The first, referred to as *Zurich region*, covers 400 km^2 around Zurich, Switzerland, for a period of 24 hours. Traffic in the area was simulated using the Multi-agent Microscopic Traffic Simulator (MMTS) [9] for the microscopic mobility, and an estimation on the daily traffic demand for the macroscopic mobility. This is the only large-scale vehicular movement trace of a metropolitan area available to date; however, the level of detail it provides is relatively low, since (i) the road map is coarse, only accounting for highways and main traffic arteries in Zurich, (ii) the MMTS is based on a queuing approach, faster but less accurate than the car-following we adopt, and (ii) the traffic demand is not as accurate as the one we dispose of, as we will later observe.

The second reference scenario is named *Zurich downtown*, and covers around 12 km^2 of the same city above, for a period of 20 minutes. The car traffic was simulated using the Generic Mobility Simulation Framework (GMSF) [23], which includes road topology information from the Swiss Geographic Information System, and car-following microscopic mobility via the Intelligent Driver Model (IDM). However, the trace does not account for a realistic traffic demand.

Also the third scenario, referred to as *Turin downtown*, is representative of a city center, covering 20 km^2 of Turin, Italy, for around 1 hour. The trace was generated using OpenStreetMap road information, as well as SUMO for the microscopic mobility. The traffic demand was built based on direct observations by the authors [24].

Traffic volumes. We first comment on the traffic present in each scenario. In the left plot of Fig. 14 we compare the traffic volume recorded in our dataset with that obtained from the Zurich region scenario. We can note that the general trend is the same, with traffic peaks between 7:00 am and 8:00
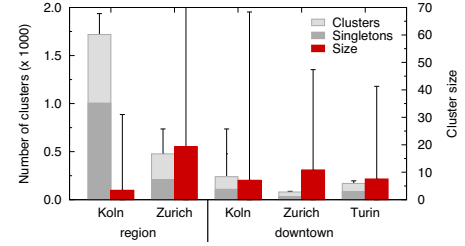
am and a reduced intensity otherwise. However, the traffic demand in the Zurich region scenario unrealistically drops to zero at around 10:00 am: we observed a similar behavior in the afternoon part of the trace, not shown here for the sake of clarity, with complete absence of traffic but in the peak hours.

The right plot of Fig. 14 shows similar information for the Zurich downtown and Turin downtown scenarios. In this case, we portray the curves with that recorded in the inner 20 km^2 of the whole Köln region only, so to perform a meaningful comparison. We refer to such a trace subset as *Köln downtown* scenario, to distinguish it from the whole dataset, termed *Köln region* from now on. From the plot, the limitations of the Zurich and Turin downtown scenarios are evident, as they can only marginally capture the demand evolution over time.

By simply looking at the traffic volumes, it is clear that large-scale mobility traces available today lack of detail, and a higher precision is paid in terms of limited road topology size and short trace timespan. Our dataset captures the best of the two worlds, providing high accuracy over a wide road topology for a whole day.

Network clustering. Fig. 15 shows the average number of clusters, i.e., groups of disconnected components in the network, as observed in each scenario. The plot also reports the mean cluster size, as well as standard deviations. By looking at the larger *region* scenarios, on the left of the plot, a clear difference emerges between our dataset and that of Zurich: the latter results in a much more connected network than the former, with vehicles grouping in less than one third of the clusters we record in our dataset. One could wonder whether that is the effect of a much higher percentage of singletons, i.e., clusters composed of one isolated node, in the TAPASCologne-based trace: from the figure, however, a similar fraction of singletons is present in both scenarios, accounting for approximately 50% of the overall clusters. The reason for such a difference is instead explained by the extremely high average and standard deviation of the cluster size in Zurich scenario: these are evidences of the existence of a few giant connected components that gather a large portion of the vehicles. Such giant components cannot instead be found in the Köln scenario, where clusters tend to be much smaller and more uniform in size.

When looking at the *downtown* scenarios, on the right side of the figure, we can note a much lower cluster number, which is consistent with the smaller size of the areas. Results are more similar throughout the different scenarios in this case, although the connectivity of the vehicular network in the Köln trace presents a significantly higher variability. The latter is clearly an effect of the fact that our dataset captures the evo-
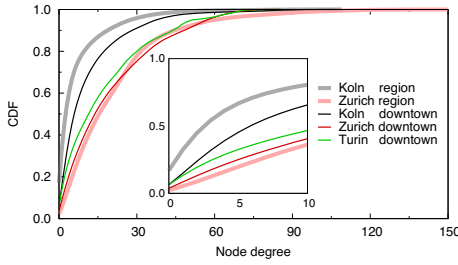
Fig. 16. Distribution of the node degree in the different scenarios

lution of the traffic over the day, whereas the other scenarios are only representative of a short time span characterized by quasi-static network clustering properties.

We can conclude that the topology of a vehicular network built on the car traffic described by our dataset is sensibly different from those obtained with currently available mobility traces. As the latter result in more connected and stable networks, one could conjecture that evaluating a network protocol or architecture with the former could lead to over-optimistic results.

Degree distribution. Fig. 16 shows the Cumulative Distribution Function (CDF) of the node degree, as recorded in the different scenarios. The inset plot provides a more detailed view of the distributions for lower values of the node degree. A significant difference can be observed between the distributions obtained from our dataset and those derived from the Zurich and Turin traces, the first been much steeper than the second. Therefore, vehicles in the Köln scenarios tend to have a smaller 1-hop neighborhoods, with only 5% of the them having more than 30 neighbors, while 60% have less than five nodes within communication range. On the contrary, in the three scenarios we take as a reference, the fraction of vehicles with large neighborhoods of more than 30 nodes grows to 25%, and only 20 to 30% of the nodes have five or less neighbors.

These results confirm those on the network clustering, and reinforce our speculation that tests conducted on mobility traces characterized by simplistic macroscopic or microscopic modeling may result in exceedingly positive performance.

VI. Conclusions and future work

In this paper, we described the generation of a large-scale urban vehicular mobility trace. The dataset is obtained by considering realistic road topology, microscopic mobility and macroscopic flows. A comparison with traces employed in the literature showed that incomplete mobility representations can lead to significantly different network topologies, possibly biasing the performance evaluation of protocols and architectures. Our dataset represents, to the best of our knowledge, the most complete vehicular mobility trace to date, and will be made available via [11]. However, we remark that we are still far from full realism: in particular, in the future we will focus on finding rigorous means to validate the dataset, on the integration of the dataset with realistic signal propagation information, as well as on a more comprehensive network connectivity analysis.

References

[1] V. Davies, "Evaluating Mobility Models within an Ad Hoc Network", MS thesis, Colorado School of Mines, 2000.

[2] F. Bai, N. Sadagopan, A. Helmy, "The IMPORTANT framework for analyzing the Impact of Mobility on Performance Of RouTing protocols for Adhoc NeTworks", *Elsevier Ad Hoc Networks*, vol.1, pp.383–403, 2003.

[3] A. K. Saha, D. B. Johnson, "Modeling Mobility for Vehicular Ad Hoc Networks", *ACM VANET*, Philadelphia, PA, USA, Oct. 2004.

[4] S. Jaap, M. Bechler, L. Wolf, "Evaluation of Routing Protocols for Vehicular Ad Hoc Networks in City Traffic Scenarios", *IEEE ITSC*, Vienna, Austria, Sep. 2005.

[5] D. Krajzewicz, G. Hertkorn, C. Rossel, P. Wagner, "SUMO (Simulation of Urban MObility): An open-source traffic simulation", *SCS MESM*, Sharjah, United Arab Emirates, Sep. 2002.

[6] M. Fiore, J. Härri, F. Filali, C. Bonnet, "Vehicular Mobility Simulation for VANETs", *SCS/IEEE ANSS*, Norfolk, VA, USA, Mar. 2007.

[7] J. Härri, F. Filali, C. Bonnet, "Mobility models for vehicular ad hoc networks: a survey and taxonomy", *IEEE Communications Surveys and Tutorials*, vol.11, no.4, pp. 19–41, Dec. 2009.

[8] H. Zhu, M. Li, Y. Zhu, L.M. Ni, "Hero: Online real-time vehicle tracking", *IEEE Transactions on Parallel and Distributed Systems*, vol.20, no.5, pp. 740-752, May 2009.

[9] N. Cetin, A. Burri, K. Nagel, "A large-scale multi-agent traffic microsimulation based on queue model," *STRC*, Ascona, Switzerland, Mar. 2003.

[10] iTetris - An Integrated Wireless and Traffic Platform for Real-Time Road Traffic Management Solutions, http://ict-itetris.eu.

[11] TAPASCologne project, http://sourceforge.net/apps/mediawiki/sumo/index.php?title=TAPASCologne.

[12] OpenStreetMap, http://www.openstreetmap.org.

[13] Osmosis, http://wiki.openstreetmap.org/wiki/Osmosis.

[14] Java OpenStreetMap Editor, http://josm.openstreetmap.de/.

[15] Simulation of Urban Mobility, http://sumo.sourceforge.net.

[16] S. Krauß, P. Wagner, C. Gawron, "Metastable States in a Microscopic Model of Traffic Flow", *Physical Review E*, vol.55, no.304, pp.55–97, May 1997.

[17] D. Krajzewicz, "Kombination von taktischen und strategischen Einflssen in einer mikroskopischen Verkehrsflusssimulation", in T. Jürgensohn, H. Kolrep (editors), *Fahrermodellierung in Wissenschaft und Wirtschaft*, VDI-Verlag, pp.104–115, Berlin-Adlershof, Germany, 2009.

[18] C. Varschen, P. Wagner, "Mikroskopische Modellierung der Personenverkehrsnachfrage auf Basis von Zeitverwendungstagebuchern", *Stadt Region Land*, vol.81, pp.63–69, 2006.

[19] G. Hertkorn, P. Wagner, "The application of microscopic activity based travel demand modelling in large scale simulations", *World Conference on Transport Research*, Istanbul, Turkey, Jul. 2004.

[20] G. Rindsfüser, J. Ansorge, H. Mühlhans, "Aktivitätenvorhaben", in K.J. Beckmann (editor), *SimVV Mobilität verstehen und lenken – zu einer integrierten quantitativen Gesamtsicht und Mikrosimulation von Verkehr*, Final report, Ministry of School, Science and Research of Nordrhein-Westfalen, Düsseldorf, Germany, 2002.

[21] M. Ehling, W. Bihler, "Zeit im Blickfeld. Ergebnisse einer repräsentativen Zeitbudgeterhebung", in K. Blanke, M. Ehling, N. Schwarz (editors), *Schriftenreihe des Bundesministeriums für Familie, Senioren, Frauen und Jugend*, vol.121, pp.237–274, Kohlhammer, Stuttgart, Germany, 1996.

[22] C. Gawron, "An Iterative Algorithm to Determine the Dynamic User Equilibrium in a Traffic Simulation Model", *International Journal of Modern Physics C*, vol.9, no.3, pp.393–407, 1998.

[23] R. Baumann, F. Legendre, P. Sommer, "Generic mobility simulation framework (GMSF)", *ACM MobilityModels*, Hong Kong, China, Apr. 2008.

[24] C. Barberis, G. Malnati, "Epidemic information diffusion in realistic vehicular network mobility scenarios," *IEEE ICUMT*, St. Petersburg, Russia, Oct. 2009.