Generation of Realistic Traces for Vehicular Mobility Simulations

Agata Grzybek  
University of Luxembourg  
Interdisciplinary Centre for Security, Reliability and Trust  
Luxembourg, Luxembourg  
agata.grzybek@uni.lu

Grégoire Danoy  
University of Luxembourg  
CSC Research Unit,  
University of Luxembourg  
Luxembourg, Luxembourg  
gregoire.danoy@uni.lu

Pascal Bouvry  
University of Luxembourg  
CSC Research Unit,  
University of Luxembourg  
Luxembourg, Luxembourg  
pascal.bouvry@uni.lu

ABSTRACT
Realistic vehicular traces are necessary in reliable VANETs simulations in order to evaluate protocols and applications designed for new technology. Because of the complexity of traffic behaviours, the generation of vehicular traces is one of the biggest challenges in VANETs research. This paper consolidates the current state-of-the art on vehicular mobility generators into three aspects: mobility models, traffic demand model and route assignment method as well as provide and overview of existing generators. The main part of the paper focuses on the improvement of traces generated by the VehiLux model over the city of Luxembourg. The quality of a large set of traces is assessed by the means of realistic traffic simulation and a proposed fitness metric. The influence of the implementation of the Gawron’s algorithm—route assignment method on the quality of generated traces is evaluated.

Categories and Subject Descriptors
C.2.1 [Network Architecture and Design]: Wireless communication; I.6.5 [Simulation and Modeling]: Model development, modeling methodologies

General Terms
Design, Experimentation, Performance

Keywords
Mobility generator, VANETs, vehicular traces, simulation

1. INTRODUCTION
Traffic congestion and its economical, social and environmental impacts are still unsolved problems of the concurrent world. Government and transportation departments have thus been recently prioritising incentives that improve mobility both on urban and highway roads. Not only investments in road infrastructure are required, but also more efficient traffic management and transportation services [24]. With the rapid development of computer and communication technologies, new opportunities and challenges appear. For instance, intelligent transportation systems (ITSs) aim to provide cost-effective and efficient solutions by using vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications [26]. Such networks of moving vehicles—VANETs (vehicular ad hoc networks) are made possible by new communications paradigms. The emergence of VANETs has raised novel research challenges different than in regular ad hoc networks due to specific mobility and connectivity patterns caused by the high speed of the vehicles and their constrained movement. VANETs are envisioned to provide reliable, secure and efficient services that have to be evaluated before can be deployed. The evaluation has to rely either on testbeds or simulations. Testbeds permit to obtain performance measures under real-word conditions but do not allow large-scale evaluations due to their cost and complexity. Realistic simulations are thus necessary in VANET development. Their quality highly depends on the degree of realism of underlying modules. On the one hand, testing of communication standards, such as IEEE 802.11p, LTE or Mobile WiMAX requires an accurate modelling of networking layers and the signals’ physical characteristics (see e.g. [10]). On the other hand, the specific mobility of nodes must also be considered, since it influences network connectivity and performance [28].

Vehicular mobility modelling produces traffic traces that specify an individual travel as a departure, destination and an exact path between them. Attempts to directly obtain traces from tracking of moving vehicles, such as taxis in San Francisco [2] or buses in Chicago [14] show many limitations, e.g. a coverage of partial traffic, a short time-span or a small geographical area. Even, when they are very accurate, real-world traces thus cannot replace artificially generated traces, for instance when there is a need to reconstruct an atypical phenomena, such as a road accident. The possibility to customise and parametrise generated traces not only makes them more flexible for evaluation purposes but also enables to predict traffic under the dynamic conditions. Thus, the current challenge in mobility modelling is the generation of realistic traffic traces at large-scale (both geographical and temporal).

In this work, we propose a method to increase the realism of artificial traces produced by the VehiLux [23] mobil-
ity model on a very large simulation area, i.e. more than 2200km$^2$. More precisely we propose to implement a route assignment technique that improves the accuracy of traces assessed by a defined fitness metric.

The paper is organised as follows. In the next section, a state-of-the-art on vehicular mobility generation and examples of existing generators are provided. Then section 3 describes the VehiLux mobility model and propose a method to increase the realism of generated traces—a route assignment technique. Section 4 presents the methodology for evaluation of traces before and after enhancement by using realistic traffic simulator and the introduced fitness metric. Finally, section 5 provides conclusions and perspectives.

2. RELATED WORK

Vehicular traces represent traffic behaviour by providing information about individual travels, i.e subsequent locations and speed values of a vehicle over time. Traces can be either real-world or artificial. Real-world traces are usually not sufficiently detailed and publicly available, thus there is a need to generate them artificially by using mobility models.

2.1 Mobility models

Mobility models reproduce vehicular movement, which is non-random, but strongly determined by the surrounding environment (i.e. road topology, traffic rules and other vehicles). It is additionally influenced by individual driver behaviour. A comprehensive study of various mobility models is given in [19]. Simple random models, like Random Way-Point or models with constrained movement (Manhattan) are not sufficient to reproduce realistic traffic mobility as shown in [15]. To better reflect vehicular movement, interactions among vehicles as well as between vehicles and the environment are described by car-interaction models that specify rules for car following, lane changing and overtaking behaviours. To increase the realism further, the complex nature of human aspect is modelled as non-deterministic behaviours. Currently existing traffic simulators, such as SUMO [9] include advanced micromobility models that can reflect both vehicular physics and driver randomness as individual traces.

2.2 Traffic demand models

Besides physical motion, vehicular mobility is also determined by a traffic demand planning, i.e. the process of choosing origins, destinations of trips and paths between them [19]. Random approaches based on stochastic choices can not reconstruct specific flow patterns visible in real-world scenarios. Vehicular traffic is not distributed evenly, but exposes dominant flows between particular areas. The prominent approach for trip planning is a 4-step traffic demand model that specifies: (a) the number of travels, (b) their origins and destinations matrix (ODM), (c) the exact path for each travel and (d) the transportation mode [20]. The model, although intuitive and simple, does not consider any time-variace of trips. The route is typically computed as the shortest path using weighted (e.g. on maximum allowed speed) Dijkstra algorithm. More behaviourally realistic activity-based models take into account underlying reasons for trips (activities taken by the travellers) [13]. An individual travel is modelled as a sequence of trips often starting and ending at a home place. The microscopic approach ensures that all decisions regarding travel choices are made at individual level which preserves the complexity in time, space and a transportation mode. Realism is further increased by modelling intra-household interactions and joint travels within one household. The activity-based methods require as input large, detailed data sets, e.g. census long forms, home interview surveys. Such models are referred as survey-based models.

2.3 Route assignment problem

In traffic networks each road segment can be presented as a link with a travel cost, e.g. the average travel time dependent on maximum allowed speed and length [6]. For each OD pair, the exact route consisting of subsequent road segment needs to be specified. A natural approach for route choice is dictated by user-selfish behaviour and the cheapest route is chosen. The shortest path can be computed by algorithms found in graph theory, like weighted Dijkstra’s algorithm, Bellman–Ford or A* (A-star) [25]. The simple application of the routing algorithms for the whole ODM results in a so called all or nothing assignment. This leads to an unnatural situation when all of the traffic for the same OD pair moves along one route. Route assignment techniques aim to distribute traffic more evenly, so the user equilibrium (UE) is achieved. In UE state, each driver choses the best route (with minimal travel cost) in respect to all other users of the road network (Wardrop’s first principle) [12]. It means that no user will benefit from the route change, and that each available alternative route for the same OD has the exact cost. Static route assignment techniques simplify the relation between travel cost and traffic flow. They assume that travel cost depends proportionally on traffic flow, but neglect the fact that traffic demand varies over the day [11]. Dynamic traffic route assignment takes into account time-dependent traffic demand and estimates travel time based on the historical and current state of a road network. Such approach enables to handle the emergence of traffic jams caused when traffic demand exceeds the capacity of a road [29]. The dynamic route assignment problem can be solved by a Gawron’s iterative algorithm [16] which is defined in following steps.

1. Assign the route for each driver by choosing the optimal route in a free (not congested) network.
2. Use simulation to calculate the actual costs for each link.
3. Recalculate the optimal routes of a certain part of the traffic using the time dependent costs from step 2.
4. Repeat from step 2 if there was a change in routes.

It is important to point out that the algorithm may not be stable and oscillate e.g. when moving a part of the traffic between two parallel routes of the same length all over again. The instability can be overcome by stochastically choice of one route from the set of alternative routes based on probability distribution.

2.4 Examples of traces generators

There exist many different approaches in generation of vehicular traces. Some of them try to derive a mathematical model from real-world measurements, like already mentioned model of buses traces in Chicago [14]. Another branch
of artificially generated traces exploits social network behaviours. For example, a model in [21] determines travel choices using communities that users visit more often and periodically revisit.

The current trend in vehicular traces generation is to combine many approaches into a single process in order to obtain the required level of realism. For example, traffic demand model can be extracted from survey statistics and tailored additionally with real-world probes. Then, individual traces can be generated by a traffic microsimulator that ensures realistic physical movement of each vehicle over an accurate map topology. Because many elements influence the mobility, they should be considered in reliable traffic generators that aim to produce realistic vehicular traces [17].

Several traces generators have been proposed in the literature. VanetMobiSim integrates realistic map topology, defines user trip sequences and implements microscopic mobility models [18]. Another project, that provides a large set of realistic vehicular traces for Zurich area in Switzerland uses statistical data to define traffic demand and a multi-agent microscopic traffic simulator to generate traces [22].

In [30], realistic traces for city of Köln, Germany (400km²) are generated by combining several elements: (i) a real map, (ii) statistical survey data that is used in (iii) microscopic activity-based traffic demand model and (iv) dynamic routes assignment. The results of SUMO simulation show that the obtained traces capture traffic evolution over a day closer to real measurements than by less sophisticated traffic generators.

The VehiLux model generates vehicular traces over even larger area of 2200km² of Luxembourg. It exploits freely accessible data sources, like open-source maps and real traffic counting data from a traffic monitoring website. Parameters of the model can be successfully discovered by using genetic algorithms (GAs) [27]. It means that model can be applied easily to any area for which map topology and basic traffic volume counting data exist. The model uses a simple Dijkstra algorithm for route assignment. In the next section we detail the process of generating traces by VehiLux model and suggest the method to improve their reality.

3. VEHILUX MODEL

The VehiLux model generates a set of vehicular traces that consider (i) spatial, (ii) temporal and (iii) behavioural aspects of traffic distribution. Uneven distribution over a space is obtained by an original traffic demand model. It specifies each route as an origin-destination pair and an exact path between them. A trace is additionally determined with a departure time that enables to reproduce the rush-hour peak in traffic demand. Traces are further assigned by a vehicle type that brings heterogeneity into microsimulation (e.g. acceleration and maximum speed capacities).

The general architecture of the model is presented on Fig. 1. The model uses freely available data: a detailed Open Street Map (OSM) and real traffic volumes at particular points obtained from the transportation department Ponts et Chaussées [1]. A single traffic volume count represents the number and type of vehicles that passed by the counter in a given time interval. The available traffic volume count locations are divided into two sets. The first set (entering points) aids a specification of traffic demand volume, while the latter (control points) a validation of the generated traces. Traffic demand is generated in two parts: the first part (outer traffic) is based on flow volumes gathered from entering points, and the second part (inner traffic) is generated from probabilistic model of attraction zones. The probabilistic model enables to select origin/destination locations by the definition of different zone types—commercial, residential and industrial, attractivity areas. Each area contains number of zones which shape is extracted from OSM maps. The probability of selecting a specific zone depends on its surface, type and attraction area that it belongs to. The exact path for the selected OD points is computed as the shortest path using a Dijkstra algorithm weighted on the travel time that is estimated as the maximum allowed speed and the length of the road. However, in real world not all drivers take the same route. We propose the application of an iterative route assignment method, namely Gawron’s algorithm, to obtain a more evenly distributed traces.

4. SIMULATION

In order to evaluate the impact of the dynamic route assignment algorithm on the quality of generated traces a simulation scenario of Luxembourgish road network was chosen. Traffic in Luxembourg exposes specific patterns of commuting flows of vehicles from neighboring countries that causes a high congestion during rush hours. In the following subsections we first present settings and new up-to-date input data for VehiLux model (see Sec. 3) that is used to generate ODM. Then, we analyse the generated set of traces both in a static and dynamic (using simulation) way. Finally, we evaluate the results of the Gawron’s algorithm.

4.1 Input data

The most recent maps were downloaded from OSM servers [3] that cover the area of 51km x 44km and contain detailed information about roads (speed limits, number of lanes, or traffic lights) and the purpose of the area usage. Fig. 2 presents data extracted from the map: the road network consists of 48006 edges of total length 5 167 670 km, 751 residential zones (green), 1788 commercial (orange) and 70 industrial (blue) of the proportional area coverage: 82%, 9%, 9% accordingly. Our solution uses the latest version of tools with carefully set parameters and is able to convert the road topology without major pathologies. Additionally, the latest real traffic counting data were collected from traffic monitoring service provided by the Ministry of
Transport [1]. VehiLux model uses 15 counters (squares) to support specification of traffic demand and 13 to evaluate generated traces (circles). The work presented in this paper uses the average value of the most recent available records from 10 following working weekdays describing a commuting traffic during the first half of a day (midnight-noon), when flows of commuters entering into Luxembourg dominate.

4.2 VehiLux results

The total set of 110,314 traces was generated for a time span of 12 hours (00:00–12:00 a.m.). Parameters of the VehiLux model were discovered by genetic algorithms as presented in [27]. For example, the inner traffic ratio was set to 0.34 that indicates the ratio of traffic generated in addition to the traffic coming from entering points. The evaluation of generated traces is a difficult task due to the lack of sufficiently detailed real-world data to compare with. For the general analysis (see Fig. 4) we reference to statistical reports of Transportation Ministry of Luxembourg [8] and [7] that provide the data about incoming traffic (see Fig. 3). It confirms that the number of generated trips as well as their distribution reflect the pattern observed in real-world. However, such an analysis can only provide an overall assessment, because reports contain mainly aggregated numbers related to a specific problem.

The quality of the obtained traces can be evaluated as a similarity with the real-world situation. We derive a quality metric—fitness, that indicates how the generated traces are consistent with real traffic volume counts [1]. Fig. 5 a) shows that the total volume of traffic passing control points reproduces closely the real-world situation. The real total flow contains 45,978 vehicles, while the generated 46,324 (only 346 vehicles more). The trend includes the peak during the morning rush hours between 7 a.m. and 9 a.m. A more detailed analysis of traffic flows at selected locations (Fig. 5b, c and d) reveals that the generated traffic is not correct at all control points. For example, at counter 403 the generated volumes are close to the reality (the total difference between generated and realistic volumes is 387), when at counter 420 traffic is too low (the difference of 1,960 vehicles) and at counter 433 too high (2,542 vehicles).

Such an evaluation is not accurate as it considers only few control points. Thus, there is a need to assess how the generated traffic demand would behave under the realistic road conditions by using a microscopic traffic simulator—SUMO 0.15. The simulator tries to reconstruct the demanded traffic flows and move vehicles on the road network in respect to physical constraints. The output of simulation can be represented on the road network as an average travel speed and number of travelling vehicles on roads in time intervals as depicted in Fig 6. The colour of a road indicates the travel speed (ranging from green for a free flow—more than 30 km/h, to red for a low speed—less than 30 km/h) and the width presents the number of cars traveling along the road. It is worth mentioning, that colouring used in this visualisation is based on the mean vehicle speed for each hour. More frequent sampling and more relative colouring would enable to produce more accurate visualisation and is envisioned as one of the future improvements.

Simulation results show that traffic congestion starts increasing at 7 a.m. The larger number of orange roads reflects the lower travel speed. As the traffic demand rises, the congestion increases during the next hours. At 8 a.m. red road segments mean the emergence of traffic jams at the most occupied locations. During the following hours, traffic jams continue to grow and less number of cars that are able to move along roads. The situation reflects the observation of traffic congestion during morning rush hours presented on Fig. 3b. However, in real world, the congestion starts resolving around 10 a.m, when traffic demand starts decreasing, but this does not happen in the simula-
4.3 Simulation of iterative route assignment

Dynamic route assignment requires to iterate over multiple simulations in order to achieve the user equilibrium. Since Gawron’s algorithm does not specify any stop condition, we decided to run 20 iterations and analysed the results.

First we study the impact of the Gawron’s algorithm on the quality of generated traces. In order to assess the accordance with the real data, we counted a total number of vehicles passing through control points during the SUMO simulation as shown on Fig. 7. We compared the results obtained for selected Gawron’s iterations (number 4, 10, 15, 20) to results for original VehiLux traces (Dijkstra). The simulation for original Dijkstra traces is able to reproduce realistic behaviour only until 7 a.m (see Fig. 5). After that, when traffic demand significantly rises, the number of counted vehicles decreases. It confirms the previous conclusion, that high traffic congestion results in lower speeds and thus smaller number of vehicles passing by control points. We can notice that at 8 a.m. the number of vehicles decreases, even in spite of the increasing traffic demand. Results obtained for Gawron’s iterations follow the trend observed in real measurements. The sudden decrease of the number of vehicles at 9 a.m. is not more present and for the 20th iteration there is even an increase. In each iteration a part of the traffic is shifted from the most congested roads to alternative routes making traffic more smooth. With every iteration more vehicles pass counters: 34 786, 35 683, 36 480 and 36 547 accordingly for for 4th, 10th, 15th and 20th iteration.

Fig. 8 presents the value of fitness computed at control points for VehiLux and simulation traces. For first four iterations the fitness value decreases as a more distributed traffic improves the fluency of the flows. The fourth iteration has the best fitness value that is 24% better than for Dijkstra results. However with the next steps the algorithm still aims in improving the fluency and moves more traffic away from still occupied control points. The decreasing traffic demand at control points results in increasing of the fitness value. In order to obtain realistic traffic behaviour, one should remember that the real-world situation is far from user equilibrium state and includes traffic jams, especially during rush hours. The volume of traffic demand on a road determines not only the emergence of traffic jams but also the capability of resolving the congestion. The challenge in generating realistic behaviour thus lies in finding a proper traffic demand that enables to both build up and relax the queue of vehicles within the required time. Consequently, to be able to mimic traffic jams, traces cannot be distributed
too evenly. It means, that we need to find a stop condition for Gawron’s algorithm at the moment where traces are the most realistic. To indicate the most realistic set of traces we use a quality metric—a fitness value and choose the iteration for which the value is the lowest, in our case—the fourth iteration.

In order to support our decision we analyse additionally the quality of obtained traces at selected control points. Fig. 9 presents measures for Dijkstra traces on 3 charts on the left side and for the fourth Gawron’s iteration on the right. Traffic on control loop number 400 was already well reproduced by VehiLux model, however original Dijkstra traces could not be consumed by simulation due to the overloaded network. Results for the Gawron’s algorithm are close to the real ones, what means that the traffic flow is not jammed. Two other control loops (420 and 443) present the situation, where traffic demand modelled by VehiLux was too high or too low. Gawron’s algorithm improves both cases, as it indicates the most occupied roads and shifts the traffic to less congested ones. Here, the improvement in VehiLux traces drives further better measurements in simulation.

Fig. 10 shows that the implementation of Gawron’s technique can improve the performance of simulation. At every step the number of vehicles arriving at their destination increases (Fig. 10a). The larger number of ended trips results in less vehicles remaining in the simulation. It means that the overall flow in the road network is more smooth. The increasing number of running vehicles at the end of simulation indicates serious traffic congestion that can not be resolved even when the traffic demand decreases (Fig. 10b). Thereby it is important to obtain this number in a declining, or at least not inclining, trend. The chosen fourth iteration has already not increasing monotonicity. Fig. 11 confirms the improvement in the performance, particularly in the average trip duration (see Fig. 11a). Although, because not all drivers use longer the shortest paths, the average trip length increases for next iterations (Fig. 11b).

Gawron’s algorithm clearly improves simulation performance by distributing traffic demand more evenly. In order to provide a general overview of results, Fig. 12 presents the visualisation of the road network for the fourth iteration. In comparison to the results for Dijkstra traces (see Fig. 6) the overall performance of the road network is better as more roads are travelled and the majority of them are green what means fluent flow. At 7 a.m. some roads, especially in the city centre, become orange and red that indicates the building up of traffic jams. The high congestion remains till 9 a.m. At 10 a.m. we can observe that traffic jams start relaxing and some red roads disappear. The desired behaviour of resolving congestion after rush hours is obtained.

The evaluation of generated traces often has to rely on visualisation tools because of the lack of numerical data. Visualisations enable to compare the obtained results roughly with driver’s empirical experience. There exists also traffic monitoring services, e.g. TomTom [4] or ViaMichelin [5], but they only provide visual snapshots of a current traffic condition and without numerical reports that unenables to perform an aggregated analysis.

### 4.4 Improving the generator

To summarise, Gawron’s technique enables to improve the realism of generated traces, as it relaxes traffic and allows to use the road infrastructure in a more efficient, thus realistic way. We rely on traffic simulation tools to evaluate the performance of the used road network and thus we have to take into account their reliability. The reliability of traffic simulation relies for instance on the correctness of the used road topology [30, 17]. Thus, the great concern should be put to obtain a bug-free map representation, otherwise a
missing road or an incorrect junction creates unpredictable bottle-necks that do not exist in real world (see Fig. 13).

Figure 13: Wrong junction in map topology.

The problems can appear both in an inaccurate map as well as during the map conversion to SUMO input format. Although, the latest versions of tools are able to produce the road topology without major pathologies, the methodology to detect map-related problems is still needed. Sources of such problems are mainly junctions and traffic lights, whose wrong representation and settings can cause deadlocks and building-up traffic jams.

5. CONCLUSIONS

This paper proposed to consolidate different aspects influencing the quality of generated traces into three design elements: a mobility model, a traffic demand model and a route assignment technique. We chose an existing traces generator—VehiLux and identified a lack of a realistic route assignment technique. We applied a Gawron’s algorithm—an iterative dynamic route assignment algorithm and evaluated traces with SUMO simulation results. The proposed solution improved the reality of obtained traces, as traffic has been distributed in a more balanced way. The efficiency of the whole network has improved, since more roads were used. The traffic demand on most congested roads was relaxed, that consequently resulted in a more fluent traffic. The work indicated the difficulties in evaluation of artificial traces, where the largest limitation is the lack of sufficiently detailed realistic traffic data. We proposed the evaluation of obtained traces with the quality metric—a fitness between the number of generated vehicles at selected control points and the publicly available real measurements. Simulation results showed that the Gawron’s algorithm improves the fitness for the first iterations but later decreases it by moving too much traffic from the selected roads. Thus, in order to obtain vehicular traces that reflect realistic traffic patterns, i.e. emerging and resolving of traffic jams, the traffic demand cannot to be distributed too evenly. We showed that traces obtained from the iteration with the best fitness is able to reproduce such a realistic traffic behaviour.

The future work envisions further improvements of the quality of generated traces. The VehiLux model can be extended with the more advanced probabilistic model of zones and areas. The atomisation of parameters tuning would enable to produce traces independently for the concerned area. Additionally, implementation of time-variant probabilities of zones and areas would allow to generate 24-hour traffic trend that reflects both morning and afternoon congestion peaks. Evaluation methods can also be improved. The more accurate fitness function would enable to indicate better the quality of generated traces. Although, in order to do that there is a need for more real world traffic data. So far, the visualisation tools can be enhanced to provide a mean for better and easier empirical evaluation of generated traces. Realistic vehicular traces generators, that are easily to implement for each scenario are an important step towards the faster development of VANETs.

6. ACKNOWLEDGMENTS

The authors would like to thank the National Research Fund of Luxembourg (FNR) for supporting the project.

7. REFERENCES


