Performance assessment of an epidemic protocol in VANET using real traces

Raul Amici ‡, Marco Bonola *, Lorenzo Bracciale*,
Pierpaolo Loreti *, Antonello Rabuffi ‡, Giuseppe Bianchi *
*DIE, Universitá di Roma "Tor Vergata", Rome, Italy
* {marco.bonola, lorenzo.bracciale, pierpaolo.loreti, giuseppe.bianchi}@uniroma2.it
‡ {raul.amici,antonello.rabuffi}@gmail.com

***Abstract*—Many vehicular ad-hoc network protocols have been validated using complex urban mobility simulators or by means of the few available real mobility traces. This work presents an extensive measurement campaign of the positions of a fleet of 370 taxi cabs that move in the city of Rome, Italy. For its street network and its traffic conditions, Rome presents a characteristic mobility pattern representative of an ancient city with heavy road congestion, providing a valuable test case to experiment VANET protocols. We exploit these traces to run a set of experiments to assess the performance of a simple epidemic protocol that we compare with the basic random waypoint model to quantify how far the performance metrics are from this baseline. The results show the possible outcomes of implementing data dissemination through an opportunistic network that uses taxi cabs as an information vector.**

I. Introduction

The recent advances in technologies [1] and in the standardization of the elements that compound the Intelligent Transportation System [2] [3] bring new attention on the theme of vehicles to vehicles communication for a smarter and safer road traffic. In this field, many researchers proposed several protocols to design and implement new Vehicular ad-hoc networks (VANET) [4] [5], usually testing them through simulations.
However, this validation methodology presents several complexities and some possible pitfalls because a plausible representation of the road conditions is highly conditioned by many time and space dependent factors, from the weather conditions to the presence of traffic lights, roads with multiple lanes, car accidents and so on.
To take in consideration all these elements, complex simulators have been build [6] and are able to accurately map the road condition once they are properly feed with the right tuning parameters.
Nevertheless this tuning is far from trivial especially for the networking researchers, so that they often use simplified models to validate VANET protocols.
For this reason, some protocols have been validated against real traces collected by monitoring the taxi cabs position in the San Francisco Bay Area and in Shenzhen [7].
This methodology exhibits the shortcoming of do not provide any flexibility degree (e.g. for varying the number of vehicles or their speed) but relieves the networking researchers from the burden of deeply characterize the urban mobility.
In this work we present a study on the performance of a simple epidemic routing protocol using real traces acquired by an extensive measurement campaign in the city of Rome: 370 taxi cabs that report their position every 7 seconds for a period of 6 months.
Due to its road topology that is far from the grid topology available in literature (figure 1), and to its traffic conditions, the evaluated scenario and the proposed traces expand the testing cases available for validating the performance of the VANET protocols. At the same time the presented traces will be proposed to be added in the Crawdad archive [7] to make them available to other researchers.
The paper is organized as follows: in section II we present the state of the art for the VANET validation strategy; in section III we describe the acquisition procedure of the traces as well as the filtering methodology to partially correct some incidental position errors typical of the trace reported via smartphones; then we characterize the statistics of the traces and finally in section we present several popular performance metrics such as the coverage and the message propagation speed, and we compare them with the popular naive random waypoint to provide a baseline.

II. Related works

Synthetic traces and mobility models for VANET

Usually the assessment of VANET protocols are performed combining two different methods for: i) generating the movements of the vehicles in the interesting area using a **mobility simulator**; ii) simulating the communication among nearby vehicles with a **network simulator** (e.g. with NS2).

In the most simply case, VANET protocols are validated using only synthetic movement traces (produced by the mobility simulator) so that cars can exchange data iff their distance is below a given threshold that depends from many factor but usually is set corresponding to the adopted technology, such as 802.11b/g/a or the more recent 802.11p [1].

A realistic mobility model should include several aspects of the real mobility (such as one way streets, traffic lights, obstacles, weather conditions, drivers behaviour etc.) that are complex to model and to take into account. For this reason, several simulators have been developed.

IMPORTANT framework [8] is one of the earlier simulator for MANET protocols. It combines several mobility models including the naive random waypoint (RWP) algorithm, RPGM model for group mobility, Freeway mobility model for considering a single high speed street and the Manhattan mobility model where vehicles move on a grid. From any of these

(a) San Francisco (b) Shenzhen (c) Rome

Fig. 1. The different road topologies for the city of San Francisco, Shenzhen, and Rome

models, the simulator can produce a connectivity graph that can be used for assessing the performances of different routing protocols. As pointed out by the authors, RWP presents several limitations because it neglects the temporal, the spatial, and the geographical dependencies of the vehicles. However it is still the "most commonly used mobility model in the MANET research community".

VanetMobiSim [9] [10] is an extension for the CANU Mobility Simulation Environment (CanuMobiSim), for vehicular mobility. VanetMobiSim can import maps from the US Census Bureau TIGER/Line database, or randomly generate them using Voronoi tesselation. This simulator goes in the direction of simulating a near-to-reality scenario, hence it offers the support for many mobility features such as multi-lane roads, separate directional flows, differentiated speed constraints and traffic signs at intersections. For this reason, it must be tuned using several complex parameters often hard to quantify.

TRANSIM [11] is an integrated set of tools developed to conduct regional transportation system analyses. It integrates several aspects of mobility, including the synthetic behaviour of public transportation and many algorithms to mimic the regional population for matching the real population demography. It uses cellular automata to simulate interactions among vehicles.

Another alternative is SUMO [12], an open source mobility simulator that uses a Gipps-model to simulate the main features of traffic flow, while taking into account a wide range of editable features such as traffic lights etc.

Due to the complexity of these simulators, often the researchers resort to simpler models such as the naive Random Waypoint, or Random Waypoint City Model [13] that includes city maps, or STRAW [14], that adds vehicular congestion and simplified traffic control mechanisms.

All these models could be enriched by adding more parameters (see [6] for a complete survey), paying the price of adding a greater complexity to the already complex environment.

In [15] authors analyse the different impact of the features introduced to simulate urban mobility and draw the conclusion that some features such as waiting at the intersections, affects the simulation results in a more significant manner than other features such as multiple lines or coordinated traffic lights.

In this paper we do not deal with the mobility model but we take the vehicular positions "as they are", analysing several long real traces acquired by tracking a fleet of taxi cabs. We compare them to the basic random waypoint not to demonstrate that the model does not fit with the reality, but to quantify how far the performance metrics are from that. Moreover, RWP, due to its simplicity, has been used in several works as a baseline [16].

Real traces for VANET

Some works use real traces [17] to study the storage capability of VANET [18] or the dynamics of network topologies [19]. Usually these real traces are provided by tracking the public transportation vehicles. Among them, the traces obtained by tracking taxi cabs are particularly important with respect to the ones obtained by tracking buses because the formers better explore the status of the city streets.

In this field, the most used public domain traces available in literature are the GPS traces of 533 taxis collected in 20 days in the San Francisco Bay area, USA [7] and the traces of 13,799 taxi cabs collected in 9 days in the city of Shenzhen. In this paper we present an analysis in the city of Rome, that exhibits a different topology if compared to the Shenzhen and San Francisco Bay area (almost) grid topology 1. Another alternative is to use some other traces available by some public services (eg. NOKIA Sports Tracker) and acquired by logging mobile phone positions. However these traces present only some segment of times and are then not suitable for an extensive analysis. Moreover it is not guaranteed that the samples in the trace refer to vehicles and not to pedestrians, train etc.

III. Evaluation scenario

In this section we describe the details of the real mobility traces and how they are acquired. Given that they are obtained by tracking the real movement of a set of taxi drivers in a city, it is particularly important to properly define the context both in terms of the acquired procedure of the location points and in terms of what is the behaviour of the drivers.

Traces acquisition methodology

Each driver has a tablet device with the Android OS and an app that sends the GPS position every 7 seconds towards a server.
On the application side, the position is updated using the *getLastKnownLocation* method of the *LocationManager* Android object and it is filtered against its precision, using the *getAccuracy()* function. This function returns the estimated accuracy in meters with 68% of probability. A sample is accepted only if the accuracy is less than 20m and it is discarded otherwise.

We subsequently filter the collected traces to mitigate some localization errors. In particular, analysing the trace we notice that there are some "oddities" that we recognize because the distance between two subsequent points is more than 125m that corresponds to a speed greater than 64km/h, a reasonable upper bound considering that the speed limit in the downtown is 50km/h. These oddities usually happen when the drivers are in some part of the city where the GPS service quality is poor (e.g. tunnels, high buildings etc.). In these cases we distinguish if the duration of the anomaly if lesser or greater than 42s (i.e. 6 points). In the former case we simply discard the "bad" samples. In the latter case (and if the anomaly is not too long, less than 8 minutes), we correct the trace by introducing artificial samples according to the short path between the endpoints of the gap. For this task we use the Open Streetmap database. Finally, if the gap is greater than 8 minutes, we consider it as a service interruption (lunch break, end of shift etc) and we take no action in this case.

Description of the sampled scenario

We limit our analysis to the center of Rome, where the density of the taxi cabs is relevant.
We consider an area of 8km x 8km whose bounds are given by the coordinates pairs (41.856, 12.442) (41.928, 12.5387).
This scenario is characterized by very thin and congested roads, high traffic volume and slow speed.
The area has been analysed using a 32x32 grid. A grid cell covers a square area of 250x250m.

Statistical parameters of the evaluated traces

We took a bigger sample of five months (October 2013 to February 2014) to derive the statistical parameters coming out from the traces. During the days, the taxi drivers can either move for serving customers (we call these movements "rides") or stay in parking areas if they are idle. For the driver speed, we analyse 1359 rides and obtain an average car speed of 12.29 km/h with a variance of 19.15. The mean waiting time is 600s. The CDF of the average speed for each ride is shown in figure 2. As we can see, there are few cabs that moves with an average speed greater than 25 km/h. This is a reasonable average speed considering the stop times (e.g. traffic lights) and the traffic congestion for the presented scenario. We point out that this value is about half of the speed limit in the area (50km/h).

Figure 3 shows the presence of the cars in the different part of the reference area. In particular, for each cell, the overlay shows the probability that there is at least one car in each cell of the grid in a reference period. If we consider

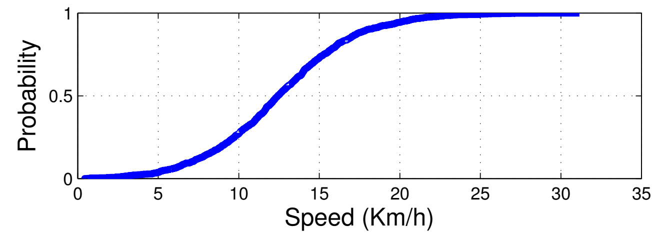


Fig. 2. CDF of the average speed for each considered ride

this period equals to 6 hours (figure 3 a)) there is an high probability that the cabs cover almost the entire area with the exemption of some zones that are not covered by the road network such as graveyards, public gardens, railway stations etc. As we consider a smaller reference period of 1 hour (figure 3 b)) the probability is greater than 0.8 only in the very center of the area and alongside the principal streets. If we shorten even more the period to 10 minutes as represented in figure 3 c) the great part of the area has a probability minor than 0.2. As a baseline result to compare with, in figure 4 we plot the same study for the case of random waypoint movement and a reference period of 10 minutes. As we can see, the area is already covered only after 10 minutes while the whole area appears as fully covered after just one hour (we omit the figure for brevity).

This study can be useful in the smart city scenario for assessing the feasibility and the performance of a metropolitan VANET where the taxi cabs carry the messages emitted by some monitoring devices (such as pollution sensors or traffic sensor etc) towards a sink node (e.g. a server of the municipality). In this way, the devices could communicate opportunistically to a cab and do not require any 3G/4G connections using the "better than nothing" network infrastructure, hence reducing the maintenance costs if the delay performance are tolerated by the specific service.

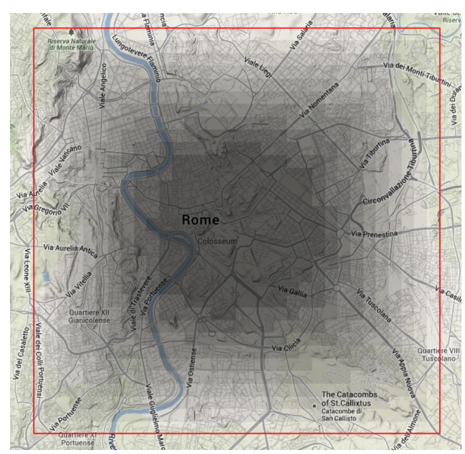
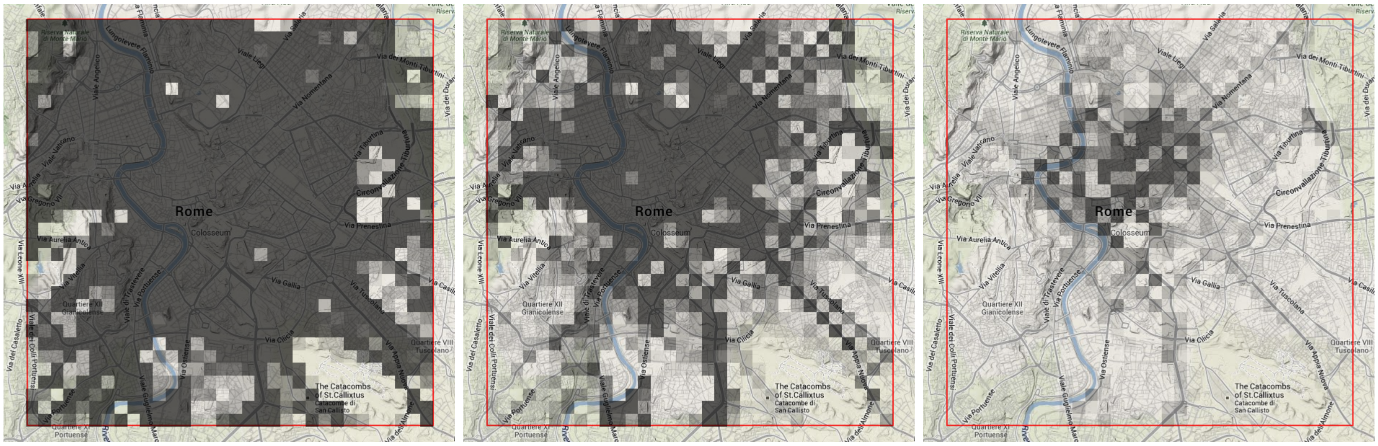


Fig. 4. Probability that at least one car enters in a cell varying the reference period using random waypoint. The grey scale overlay represents the probability from 0 to 1 with step of 0.1: darken area means a greater probability

Discussion

Rome is currently the city with the highest number of transportation on wheels in Italy. Its road topology is very



(a) 6 hours (b) 1 hour (c) 10 minutes

Fig. 3. Probability that at least one car enters in a cell varying the reference period using the real traces. The grey scale overlay represents the probability from 0 to 1 with step of 0.1: darken area means a greater probability

irregular and very far from the grid topology.

We point out that there are some streets dedicated only to public transportation (bus and taxi), therefore the presented results can not be applied to general user mobility case. Moreover taxi drivers has a different behaviour from other citizens because their movements depend on the customers demand.

In general this demand varies during the week and across different months, given that there is an higher demand in the summer because the tourism increases. For this reason we limit the analysis in section IV to just one month, from February 1st 2014 to March 2th 2014. Moreover there could be sensible variations depending on holidays, weather conditions, special events etc.

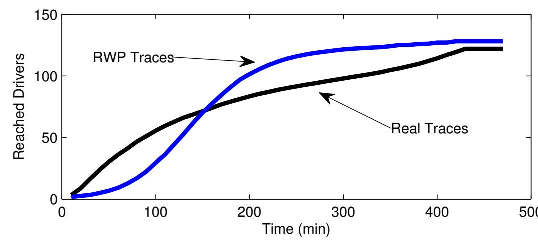
Finally there are some municipality regulations that impose to give priority in rides assignment to the drivers parked in taxi parking lots. This leads to a periodic behaviour in which the drivers start from a parking lot, serve the request and then go back to the parking lot.

IV. VANET APPLICATIONS

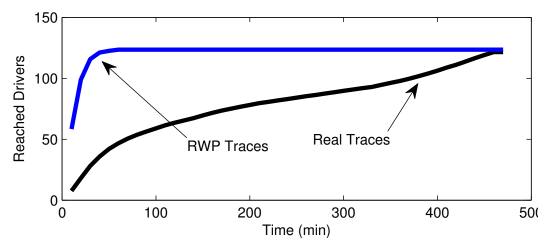
In this section we evaluate a simple epidemic routing protocol on the collected real traces. The protocol works as follows: an initial message spawns on an uniformly distributed random car at a given time. That car is called "infected". Then the message is passed to every not infected car whose position is within a given range (*coverage range*) from the infected one. Once a car receives the message, it becomes infected and participates in the data dissemination.

It is evident that after a certain period every car will receive the message hence the steady state of the process is not interesting. For this reason we present the transitory part of the diffusion process. We remark that this data dissemination (often called "epidemic routing" in literature) is the fastest one available in the Delay Tolerant Networking field. Even if this routing presents several disadvantages such as the memory occupancy and the huge number of data circulating in the network, it can serve as an upper bound for any other routing scheme.

For all the presented results we analyse 30 days of movement for about 370 taxi. For each day we conduct four different simulations in which we vary the origin of the messages. Every



(a) Coverage range 50m

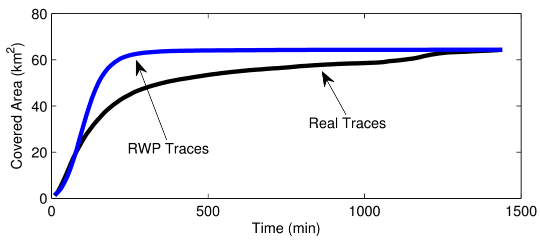


(b) Coverage range 250m

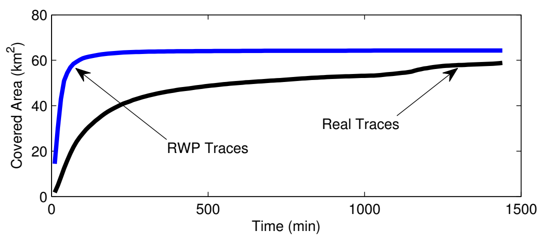
Fig. 5. Number of infected drivers as a function of the time.

simulation starts at a random time during the interval 0 am to 6 pm. We plot the average results together with the ones obtained through cars that move according to a simple random waypoint algorithm in which we set the speed to a constant value equals to 12.29 km/h and the wait time has been set to 600s. These values are respectively the values of the average speed and the average wait time calculated on the real traces.

Figure 5 shows the number of reached (infected) vehicles as a function of the time comparing the real traces with the one produced by the random waypoint and varying the coverage range from 50m (figure 5 a)) to 250m (figure 5 b)). In figure 5 a) the infection process on the real traces presents an higher speed in the initial part. The reason of this behaviour can be accounted to the distribution of the drivers in the area, that is more concentrated in the center for the real case with respect to the random waypoint case, as we can see by comparing figures 3 c) with 4. If we increase the range, from 50m to 250m, the RWP presents a more rapid diffusion due to the independence



(a) Coverage range 50m



(b) Coverage range 250m

Fig. 6. The covered area as a function of the time.

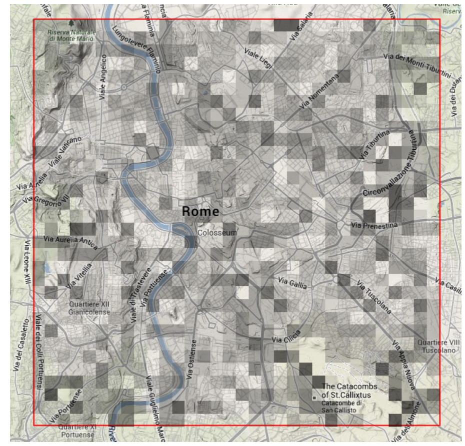
assumption of the paths that are typical unrealistic. Despite several literature works present a coverage range of 250m or more, we point out that for 2.4 and 5 GHz communications, in the proposed urban scenario, 50m seems more realistic because of the presence of several obstacles and a very high level of interference.

Figure 6 shows the covered area defined as the sum of the areas of the cells where there is at least one infected driver before a given time. The covered area saturate to the maximum area , $64km^2$. For the considered scenario, while the RWP exhibits a relevant boost if we increase the coverage range, in the real case the difference is negligible.

In figures 7 and 8 we show the average message propagation speed defined as the time elapsed from the instant when a message enters for the first time in a cell and when it will be passed to the first neighbour uninfected cell. This measure is important to understand the gradient of the message propagation speed. As shown in figure 7, in the real case the speed values are scattered in the whole area and they are not significantly affected by the increase of the communication radius range. In the random waypoint case represented in figure 8, the speeds are more uniform and, as expected, they present greater values in the center of the area, and this is exacerbated if we vary the communication range from 50m to 250m.

V. Conclusions

In this paper we presented the results of an extensive measurement campaign where we collect the position of 370 taxi cabs for 6 months working in the center of Rome, characterized by high congestion streets and a road topology that is far from the grid. We exploited the real traces to show some performance metrics related to the epidemic diffusion of a message that can be considered as an upper bound for the message propagation in delay tolerant and vehicular area networks. All the results are compared with the random waypoint case to provide a baseline. Given the characteristic of the acquired traces, all the results are deeply influenced by



(a) Coverage range 50m

(b) Coverage range 250m

Fig. 7. Average message propagation speed for real traces

the major concentration of the drivers in the center where they can find more work, even if they also cover all the peripheral areas, but with a minor density. The presented results and traces could be used for supporting and validating emerging services in the smart city field or to assess the performances of VANET protocols in the presented scenario that, to the best of our knowledge, has not been reported yet in the literature.

References

[1] C. Campolo and A. Molinaro, "On vehicle-to-roadside communications in 802.11p/wave vanets," in *Wireless Communications and Networking Conference (WCNC), 2011 IEEE*, March 2011, pp. 1010–1015.

[2] E. T. S. Institute, "Etsi es 202 663 v1.1.0," *Transportation Research Part C: Emerging Technologies*, 2009-11.

[3] C. to car communication consortium. [Online]. Available: http://www.car-to-car.org/

[4] F. Li and Y. Wang, "Routing in vehicular ad hoc networks: A survey," *Vehicular Technology Magazine, IEEE*, vol. 2, no. 2, pp. 12–22, 2007.

[5] M. Altayeb and I. Mahgoub, "A survey of vehicular ad hoc networks routing protocols," *International Journal of Innovation and Applied Studies*, vol. 3, no. 3, pp. 829–846, 2013.

[6] E. Spaho, L. Barolli, G. Mino, F. Xhafa, and V. Kolici, "Vanet simulators: A survey on mobility and routing protocols," in *Broadband and Wireless Computing, Communication and Applications (BWCCA), 2011 International Conference on*, Oct 2011, pp. 1–10.

[7] [Online]. Available: http://crawdad.cs.dartmouth.edu/

(a) Coverage range 50m

(b) Coverage range 250m

Fig. 8. Average message propagation speed for the random waypoint case

[8] F. Bai, N. Sadagopan, and A. Helmy, "The important framework for analyzing the impact of mobility on performance of routing protocols for adhoc networks," *AdHoc Networks Journal*, vol. 1, pp. 383–403, 2003.

[9] J. Härri, M. Fiore, F. Filali, and C. Bonnet. (1999) Vanetmobisim. [Online]. Available: http://vanet.eurecom.fr/

[10] J. Härri, F. Filali, C. Bonnet, and M. Fiore, "Vanetmobisim: Generating realistic mobility patterns for vanets," in *Proceedings of the 3rd International Workshop on Vehicular Ad Hoc Networks*, ser. VANET '06. New York, NY, USA: ACM, 2006, pp. 96–97. [Online]. Available: http://doi.acm.org/10.1145/1161064.1161084

[11] L. Smith, R. Beckman, D. Anson, K. Nagel, and M. Williams, "TRANSIMS: Transportation analysis and simulation system," in *Conference: 5. National transportation planning methods applications conference, Seattle, WA (United States), 17-21 Apr 1995*, 1995.

[12] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of SUMO - Simulation of Urban MObility," *International Journal On Advances in Systems and Measurements*, vol. 5, no. 3&4, pp. 128–138, December 2012. [Online]. Available: http://elib.dlr.de/80483/

[13] J. Kraaier and U. Killat, "-the random waypoint city model-: user distribution in a street-based mobility model for wireless network simulations," in *Proceedings of the 3rd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*. ACM, 2005, pp. 100–103.

[14] D. R. Choffnes and F. E. Bustamante, "An integrated mobility and traffic model for vehicular wireless networks," in *Proceedings of the 2nd ACM international workshop on Vehicular ad hoc networks*. ACM, 2005, pp. 69–78.

[15] A. Mahajan, N. Potnis, K. Gopalan, and A.-I. A. Wang, "Urban mobility models for vanets," in *IN PROC. OF 2ND WORKSHOP ON NEXT GENERATION WIRELESS NETWORKS*, 2006.

[16] A. Mahajan, N. Potnis, K. Gopalan, and A. Wang, "Evaluation of mobility models for vehicular ad-hoc network simulations," in *IEEE International Workshop on Next Generation Wireless Networks (WoN-GeN)*, 2006.

[17] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "A parsimonious model of mobile partitioned networks with clustering," in *Communication Systems and Networks and Workshops, 2009. COMSNETS 2009. First International*, Jan 2009, pp. 1–10.

[18] B. Liu, B. Khorashadi, D. Ghosal, C.-N. Chuah, and H. M. Zhang, "Analysis of the information storage capability of vanet for highway and city traffic," *Transportation Research Part C: Emerging Technologies*, vol. 23, pp. 68–84, 2012.

[19] Y. Chen, M. Xu, Y. Gu, P. Li, and X. Cheng, "Understanding topology evolving of {VANETs} from taxi traces," *Advanced Science and Technology Letters*, vol. 42, no. Mobile and Wireless, pp. 13–17, 2013.