Analysis of urban traffic data sets for VANETs simulations

Alice Castellano* and Francesca Cuomo§

*University of Roma "Sapienza", Rome, Italy Email: alice.alice@alice.it
§DIET at University of Roma "Sapienza", Rome, Italy Email: francesca.cuomo@diet.uniroma1.it

I. INTRODUCTION

Vehicular Ad-hoc Networks (VANET) are self-organized, distributed communication networks built up from moving vehicles where each node is characterized by variable speed, strict limits of freedom in movement patterns and a variety of traffic dynamics. In the last few years vehicular traffic is attracting a growing attention from both industry and research, due to the importance of the related applications, ranging from traffic control to road safety. However, less attention has been paid to the modeling of realistic user mobility through real empirical data that would allow to develop much more effective communication and networking schemes. The aim of this study is then twofold: on one hand we derived real mobility patterns to be used in a VANET simulator and on the other hand we simulated the VANET data dissemination achieved with different broadcast protocols in our real traffic setting.

II. ANALYSIS OF GPS DATA TRACES

In order to achieve a realistic modeling for VANETs, we analyzed 1.17 GB of raw data in Microsoft Excel sheet format containing GPS traces from more than 50,220 vehicles. These GPS traces refer to 2% of total vehicles during the month of May 2010 measured in the Grande Raccordo Anulare (GRA) area, the circular highway surrounding the city of Rome. Each vehicle sends its GPS trace record every 30 seconds. A variety of information was captured in each record, including the record ID, vehicle ID, geographical coordinates, speed, quality of GPS signal. The first step of our analysis was based on the segmentation of the GRA area in 29 different segments of length \(L_j\), \(j = 1, \ldots, 29\), where the main exits on the GRA highway are the starting and ending points of each segment. Then vehicles were divided in two sets according to their traffic direction: clockwise and counterclockwise. This last step allowed us to calculate the distance of each vehicle from the closest preceding one. Four different time periods of four hours each have been considered: 1.) [7a.m-11a.m.), 2.) [11a.m.-3p.m.), 3.) [3p.m.-7p.m.) and 4.) [7p.m.-11p.m.). Inter-vehicle distance distribution and speed distribution were obtained for each of the four time periods. This analysis showed that the highest density of vehicles is in the time period between 3p.m. and 7p.m., which has the largest number of detected vehicles (9732 vehicles). In this range of time we also have the largest number of vehicles with inter-distance \(< 50\) m and the lowest average speed (64.4 km/h) compared with the other three time periods considered.

On the contrary, as we expected, the sparsest time period is the one between 7p.m. and 11p.m. with the lowest number of vehicles (978), largest inter-vehicle distances and highest average speed (95.5 km/h).

III. INJECTION OF REAL DATA TRACES IN THE SIMULATIONS AND PERFORMANCE ANALYSIS

To set up a realistic mobility simulation of the urban traffic in the GRA area the available data have been re-densified to have a realistic amount of vehicles (100%). The mathematical model used for densification is briefly described by the following relations, where \(h\) is the sampling temporal interval (30 sec in our study), \(v_i\) is the average speed of vehicle \(i\), \(m_j\) is the number of the estimated traveling vehicles in the \(j\) segment, \(n_j\) is the number of detected GPS signals in the segment \(j\), \(L_j\) is the length of the \(j\) segment, \(a\) is the probe vehicles penetration rate (2% in our study) and \(q_j\) is the estimated flow on the segment \(j\). We define \(d_i = v_i \times h\) and we have on the segment \(j\):

\[
m_j = \sum_{i=1}^{n_j} \frac{d_i}{L_j}
\]
and the resulting flow as \( q_j = \frac{m_j}{a} \).

The above densification rules have been applied to the traffic flows in the time period with the highest density, 3-7p.m. and have been used to inject on each segment \( j \) and amount of vehicles equal to \( q_j \). It is to be notices that vehicles are injected and extracted in correspondence to each of the GRA exits. Then the generated flows have moved in the GRA scenario by using SUMO (Simulation Of Urban Mobility) [3] and then used to run communication protocols by using NS-2 (Simulator Network 2) [4] and MOVE (Mobility model generator for Vehicular Networks) [5] (see Fig. 1). These tools provide, respectively, vehicles mobility, simulation of network protocols and a graphical interface between networking and mobility. The protocols simulated to provide message spreading in the proposed scenario are:

- Flooding: simply floods the network with messages.
- DBF: based on DDT (Distance Defer Transmission) [6], a node \( A \) receiving a message from node \( B \) waits \( T_{\text{max}}(1 - \frac{d_{AB}}{R_{\text{max}}}) \) before forwarding the packet, where \( R_{\text{max}} \) is an upper bound of the radio range and \( d_{AB} \) is the Euclidean distance between \( A \) and \( B \).
- DBF hop count: enhanced version of DBF, avoids erroneous propagation interruptions that in standards DBF are noticed when two forwarders are so close that their timers are almost the same, so they forward the same message, which triggers the inhibition rule at the next hop, thus stopping propagation.
- RND: a random selection of the forwarder where each node receiving a message sets a timer to a value between 0 and 100 ms before forwarding and when a forward takes place this inhibits neighbor nodes.

Simulations last 100s and were performed with a bit-rate of 160 kbps, placing a single RSU (Road Side Unit) near an Exit (#19), on the southern part of the GRA highway. The metrics used to compare performances of the different protocols are: MAC Collisions, PDR (Packet Delivery Ratio, average fraction of messages received by a node), Average End-to-end delay (average time needed for a message to reach the farthest node from the RSU). Flooding protocol achieves a good performance only when few nodes are present but behaves really bad in presence of more traffic. As we expected, it has the highest number of collisions. DBF hop-count achieves the best results compared to the others, avoiding the packets to stop. The graphs in Fig. 4 show that DBF hop-count obtains the highest PDR value, low number of Collisions and an acceptable value of Average End-to-end delay.

ACKNOWLEDGMENTS
Authors are grateful to Mario De Felice for the work done in the NS-2 simulations set up.

REFERENCES
[1] GPS traces from “Progetto Pegasus, Industria 2015,” May 2010.
[2] Mathematical model from “Progetto ENEA Smart Cities' by G. Fusco, C. Colombaroni, A. Gemma, G. Ciccarelli, S. Lo Sardo, Sapienza Universitat di Roma, 2011.
[3] D. Krajzewicz and C. Rossel, “Simulation of Urban Mobility (SUMO), German Aerospace Centre, 2007, available at: http://sumo.sourceforge.net/index.shtml
[4] K. Fall and K. Varadhan, “ns notes and documents, The VINT Project,” UC Berkeley, LBL, USC/ISI, and Xerox PARC, February 2000, available at: http://www.isi.edu/nsnam/ns/nsdocumentation.html
[5] “MOVE (MObility model generator for VEHicular networks): Rapid Generation of Realistic Simulation for VANET,” 2007, available at: http://lens1.csie.ncku.edu.tw/MOVE/index.htm
[6] P. Salvo, F. Cuomo, A. Baiocchi, and A. Bragagnini, “Road side unit coverage extension for data dissemination in vanets,” in Wireless On-demand Network Systems and Services (WONS), 2012 9th Annual Conference on. IEEE, 2012, pp. 47-50.
[7] P. Salvo, M. De Felice, F. Cuomo, A. Baiocchi “Infotainment traffic flow dissemination in an urban VANET,” IEEE Globecom 2012.