Adv Syst Sci Appl 2020; **04**:36–44
Published online at https://ijassa.ipu.ru.

Statistical Properties of VANET-based Information Spreading

Imre Varga[1]*, Gergely Kocsis[1]

[1] *University of Debrecen, Department of IT Systems and Networks, Debrecen, Hungary*

Abstract: In this work, we present a robust model of vehicular ad hoc networks (VANET) in order to study information spreading on such topologies. Vehicles are moving along the fastest routes between their starting points and their destinations on a map derived from a real urban topology. Vehicles can exchange information through the use of short-range wireless communication devices. The source of information is a roadside unit which provides packets for the nearby vehicles. These agents can carry and forward messages for others. As a result of the data dissemination, the major part of the system becomes informed quite quickly. Presented results include the investigation of information spreading in the system, e.g. the time evolution of the average awareness, the age distribution of information owned by separate vehicles and the statistical properties of time intervals between information exchange events. It has been shown how the effectiveness of this complex system depends on the density of intelligent devices. Scale-free behavior was found by time series analysis. Our computer simulation results can help to design smartcity applications of the future.

Keywords: agent-based simulation, VANET, information diffusion, traffic simulation, time series analysis

1. INTRODUCTION

The spreading of information in vehicular networks plays a key role in many smart city services. Because of this, the topic is in the focus of research in the last decade. The aim of these applications is to make urban traffic safer and comfortable. Previous studies have analyzed the topological properties of urban road maps [1, 2] and the traffic flow was measured and studied [3] in some other works in order to increase the efficiency of these intelligent transportation systems. Several different algorithms and methods were developed to simulate the motion of vehicles and generate traffic in urban or in highway environment [4–6]. The communication of moving wireless devices (possibly carried by members of these vehicular networks) may be described by using standardized communication protocols, like Dedicated Short Range Communication (DSRC) or IEEE 802.11p standard [7–9]. In VANETs (Vehicular Ad hoc NETwork) both the routing [10–12] and the broadcasting [4, 13] are actively investigated fields.

Distribution of information was also studied in different wireless systems, e.g. in mobile peer-to-peer systems [14] or self-organized sensor networks [15]. Some questions related to the statistical properties of the general spreading processes in VANETs are however still open. The goal of our research is to introduce and analyze a new framework in order to be able to answer some of the following questions. What are the limits of the information spreading? Can we reach all actors of the traffic system based only on self-organization? Do all vehicles own up-to-date information? Similar questions have been appeared and already answered in

*Corresponding author: varga.imre@inf.unideb.hu

social networks [16,17], but due to the continuously changing topology, the characteristics of spreading can be very different.

In Section 2 the construction of the realistic urban topology is shown. Section 3 presents the details of the simulation of vehicular motion. Information spreading based on carry-and-forward and multihop broadcast dissemination schemes is presented in Section 4 and then the first results of our studies are shown in Section 5. At the end we close by some conclusions.

2. UNDERLYING MAP TOPOLOGY OF SIMULATIONS

We applied a real city map, namely the map of our home city as the underlying network topology to reach a realistic simulation environment. The dataset describing the map was gained and is available at the page of the OpenStreetMap project [18]. Since in this work we would like to use this map for agent-based simulation a much more simplified topology is needed. Because of this the source was reduced keeping the topology of the crossroad network and the distances between junctions, but losing the real geographical locations of road sections.

According to the original `osm` format any crooked road can be built up from shorter straight segments and the geographical coordinates of their endpoints. In this way, a road section between crossroads can be described by a list of internal nodes with degree 2. In our approach, the shape of a road section is negligible and only the length of the section is important. This was the base of our topology simplifying method. In case of any two road segments between nodes $A - B$ and nodes $B - C$, node B was eliminated if it has no other neighbors than A and C, merging the segments to only one longer segment between nodes $A - C$ with a distance equal to the sum of lengths of the previous two segments.

After the above reduction process the resulting network can be analyzed from two different points of view. (i) Taking a look to the network as an abstract topology (undirected graph) we can find that there are 3422 nodes (junctions) connected by 4812 links (road sections). It was found that 84% of these nodes have a degree of 3 or 4. In the unit of link number (ignoring road length) the diameter of the network is 96. (ii) From the geographical aspect however our network still has some spatial properties. Taking into account the distances, the average distance between two crossroads is $121.5m$, however, the distribution is quite wide, there are almost 3 orders of magnitude difference between the shortest and the longest road section. The average distance between two randomly chosen nodes is $4.1 \pm 1.9km$. (More details are available in [19, 20].)

3. MOTION OF VEHICLES

On the above described map vehicles are moving from their randomly chosen starting node toward their randomly chosen destination node along the fastest path. Even though today many different navigation options are available to be considered, practice shows that drivers usually use a route with the shortest travel time instead of the shortest distance route, or the smallest number of left turns [21]. The original data set contains information about the $rank$ of all road segments (e.g.: primary, secondary, residential, living street, etc.). The average speeds of cars depend on the rank of the road. Based on the speed prediction/offer of the Google Maps [22], different average speed is applied in case of different road ranks. Thus the shortest and the fastest route can be different.

Vehicles move with constant speed between two neighboring nodes, at a crossroad they turn according to their route (and perhaps change speed). Traffic jams, traffic lights or the finite size of vehicles are not taken into account during the simulation because from the point of the later spreading process the short-term fluctuations of the speed of cars are negligible. Thus, in this kind of mean-field approach, the velocity of vehicles is not influenced by the actual traffic situations. Even though, the source and destination nodes are random the density of the traffic is really diverse due to the topology (connectivity, ranks).

We assumed that the number of moving cars in the system at a given time can be constant since the simulated (typically few 10 minutes long) time interval is small compared to the daily life cycle of a city or the duration of rush-hour traffic. In this way, different scenarios (e.g. rush-hours or off-peak time) can be simulated separately using a distinct number of cars. At the beginning of the simulation, there are N vehicles in the system. Later, when a vehicle arrives to its destination, it is removed and immediately a new one is initialized and started. At the beginning of the simulation, all the cars are just departed. It is easy to understand that in order to avoid artificial transient effects the measurement related to spreading is started only later ($t = 0$) when the system becomes randomized, however, the simulation of the traffic is started at $t = -t_0$. The length of the randomization time interval ($-t_0 \leq t < 0$) is longer then most of the trips ($t_0 = 750s$, average travel time is $459 \pm 261s$), so when the scientific observation is started all the initial cars have been arrived and others are launched in different time moments. The simulation is stopped at $t = t_{max}$. The system evolves in discrete time steps. The time step Δt is small enough to move only a few meters, so it is tiny compared to the whole simulation time $\Delta t \ll t_0 + t_{max}$. The time interval of the analysis ($0 \leq t \leq t_{max}$) is long enough to cover several generations of vehicles. We found that in order to get reliable results the total number of simulated cars (N_t) has to be at least five times greater than number of cars at a given moment ($N_t > 5N$).

This approach of macroscopic traffic is quite simplified. The motion of vehicles is much slower than, spreading of information (detailed in the next section) due to the consecutive quick message forwarding. This fact allows us to neglect more details of the micro-level traffic. In addition to the approximate spatial distribution of agents, motion is only required to regularly change the communication topology.

4. SPREADING OF INFORMATION

In the system, smart vehicles are represented by agents able to interact by short-range communication. If the distance of two vehicles at a given time moment is less than the range R of the wireless communication, they can exchange information. If agent i can receive information from agent j, communication to the opposite direction is also possible technically. Based on this, in our model the agents can have two different states. On the one hand agent i can be $uninformed$, so it has not received any data (denoted by $S_i = 0$). On the other hand, it can be $informed$, so it has already got some data (denoted by $S_i = 1$). Beside this vehicle-to-vehicle (V2V) communication, there is infrastructure-to-vehicle communication (I2V) as well. In the latter case the On Board Units (OBU) of smart vehicles can receive information from Road Side Units (RSU).

In our model initially all agents are in uninformed state and only one RSU is present, playing the role of an information source. When an agent passes close enough to the RSU it receives new up-to-date public information (e.g. traffic or weather alert). The actual content of messages is negligible. The agent stores it together with the actual timestamp and later it shares with others within the communication range. If one of these neighboring agents is uninformed it becomes informed. If both agents that are in contact have been already informed, the agent with an older timestamp will update its knowledge storing the newer information with its given timestamp. Thus information can spread in this dynamically changing network from the RSU to any vehicle even if they have never passed by the RSU. In order to characterize agent i in detail we introduce the quantity T_i which is the latest/newest timestamp of information owned by the informed agent i or $T_i = -1$ if agent i is uninformed. (So $T_i > 0$ is the simulation time when the given information unit carried by agent i entered into the system by the RSU.) The behavior of the system is shown in Fig. 4.1. A vehicle (agent i) proceeds from node A to node D. It goes by the RSU in node B receiving new information at $t = T$. An other vehicle (agent j) moves from node E towards node F. Both of them are

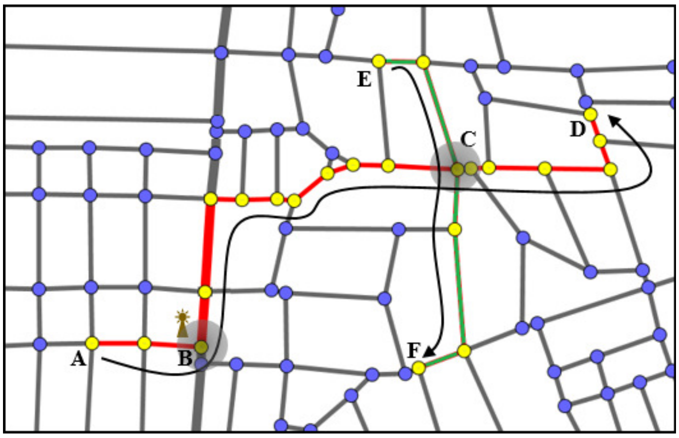

Fig. 4.1. The behavior of the system.

in the vicinity of the node C at the same time. Since they are within the range R, agent i can transmits the information to agent j. Between nodes A and B agent i is uninformed, but between B and D it is in an informed state, having timestamp T. Agent j becomes informed at node C and possesses also timestamp T between nodes C and F.

At simulation time t, an informed agent i has information with age $A_i = t - T_i$. The average age of information $\langle A \rangle$ owned by agents in a given time moment can be written as

$$\langle A \rangle = \frac{\sum_i T_i S_i}{N^{inf}}, \tag{4.1}$$

where N^{inf} is the number of informed agents, defined as $N^{inf} = \sum_i S_i$. A large value of N^{inf}/N indicates extensive information spreading. When the average age of information $\langle A \rangle$ is low, it means that our smart traffic system is in an up-to-date phase. Thus the number of informed agents N^{inf} and the average age of information $\langle A \rangle$ are good measures of the effectiveness of information spreading in VANET.

5. RESULTS

Since during the simulation an SI (Susceptible-Infected) model [23] is applied, initially more and more agents become informed. However, the system never reaches a fully informed state, because during the simulation new, uninformed agents appear in the system, while informed ones disappear as they reach their destination. Investigating the time evolution of the agents it was found that the system reaches a steady state described by saturating functions. In Fig. 5.1 a, one can observe that at $t = 0$ (when the RSU is just activated) there are no informed agents in the system, but soon some agents pass by the information source of the infrastructure. Then the vehicles carry the information during their motion to different places of the city meanwhile they behave as secondary information sources speeding up the spreading of information so leading to increasing $N^{inf}(t)/N$ function with a significant slope.

After a quite short time period, spreading slows down resulting in saturation of the number of informed agents. The average movement of vehicles during a simulation step Δt is the half of the applied range of communication R. (Of course, increasing range R speeds up the spreading.) The reason why the system reaches this almost steady state is the fact that the propagation of information can be faster than the motion of vehicles. The saturation level depends on the number of agents (the density of smart vehicles in the city) and in most cases the $N^{inf}(t)/N$ curves never reach 1.0. This is shown in Fig. 5.1 b. As one can observe the information coverage of VANET can be effective only if the number of smart vehicles

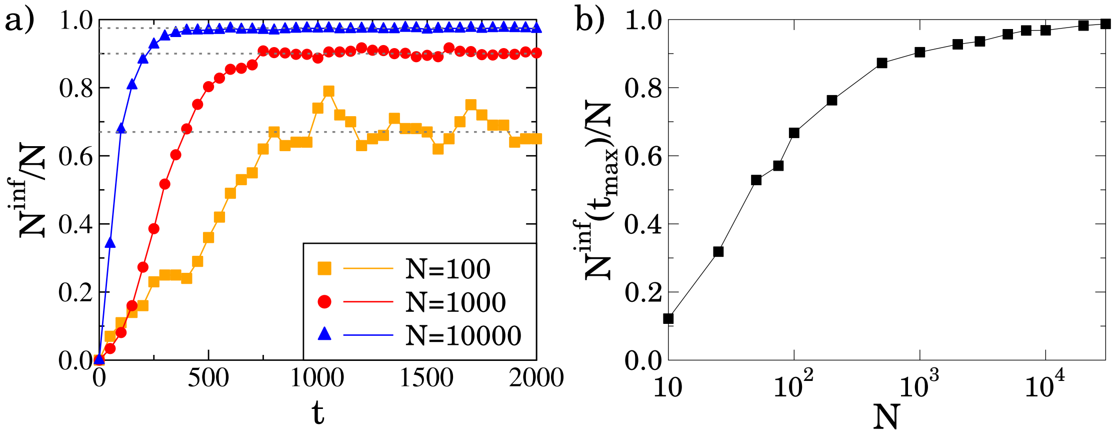

Fig. 5.1. a) Number of informed agents (vehicles) as a function of time for different numbers of agents. After a short time period a saturation is achieved at a quite high value. b) The previous saturation level depends on the number of vehicles in the system (of course more smart vehicle leads to higher level of awareness).

exceeds a given threshold (about few hundreds of vehicles in the case of the medium-sized city Debrecen).

Even though in most cases the number of informed agents N^{inf} is proved to be relatively high in the system, the really important questions are the following ones. How old is the average information? Is the system in an up-to-date phase continuously? The average information age as a function of time $\langle A\rangle(t)$ can give the answers. As it is shown on Fig. 5.2 a, most of the agents have relatively young information. Recent information from RSU overwrites the system very quickly without any outer control. Of course the saturation level of $\langle A\rangle(t)$ (far from the opening time period) is determined by the number of agents. More smart vehicles lead to a more up-to-date system. (See the Fig. 5.2 b.) The average age of information is even less than the length of the time period needed to reach the saturation of the number of informed agents.

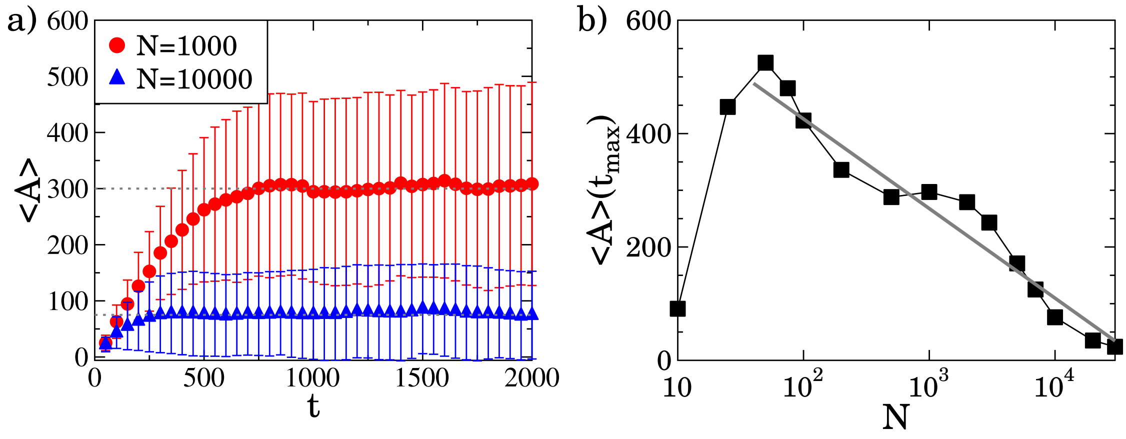

Fig. 5.2. a) The average age of information owned by the vehicles as a function of time. It shows saturation for different system size. b) The average age of information in the saturation phase decreases logarithmically with the number of vehicles, so a denser vehicle park in the city results in a more up-to-date system.

The above averages describe the system on a macro scale. For microscale characterization, we analyzed the time intervals between information exchanges for all agents. The average information age can be low only if the time interval Δt_{recv} between two subsequent information receive events is short for each vehicle. As one can see in Fig. 5.3 a, the

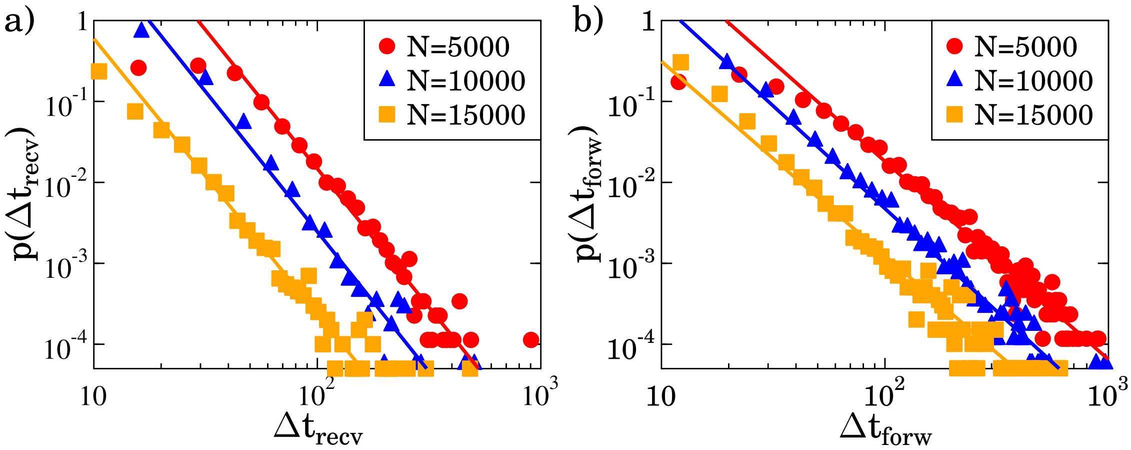

Fig. 5.3. The distribution of the average message receiving (a) and forwarding (b) time interval of vehicles. They obey a power law with exponents around 3.4 (a) and 2.5 (b) respectively, independently of the vehicle density.

distribution of Δt_{recv} has a power-law form: $p(\Delta t_{recv}) \sim \Delta t_{recv}{}^{-\gamma_{recv}}$. This means that most of the agents frequently receive new information packets, but some of them keep old information for relatively long time. The density of vehicles has no effect on the value of the exponent which is close to $\gamma_{recv} \approx 3.4$. The distribution of the time intervals between message forwards Δt_{forw} also proved to be scale-free, however its exponent γ_{forw} is definitely lower, it is around 2.5. (See Fig. 5.3 b.) One can ask why this difference in the distributions appear even though the number of receive and forward events are the same. This can be explained however by conditions of information exchange. When vehicles approach each other within the communication range, only the up-to-date agent can forward its recent information. In this way, an agent, who has a quite old information packet has to wait a long time to get a chance to forward it. Contrarily old information can be overwritten by almost any other message, so the high receiving chance leads to very few old information to overwrite.

On micro scale, one can follow the spreading of each information holding any given timestamp T. Agents receive new information from the RSU at time $t = T$. Then it is transmitted to other vehicles, so the number of agents N^T having the same timestamp T is increasing. Sooner or later they will be updated so $N^T(t)$ is decreasing, finally all of these information units will disappear at $t = T + T_L$, where T_L denotes the lifetime of the given information. Note that T_L is not the lifetime of a given information packet, but the time interval when the given information (with a given timestamp) is present somewhere in the system. The time evolution of $N^T(t)$ is qualitatively similar for all timestamps, however quantitatively they are very different. The general form of the $N^T(t)$ function can be obtained by rescaling and averaging all the separate curves. The result is illustrated in Fig. 5.4 a. It shows that at the beginning information spreads very quickly but after T_E time the expansion reaches its maximum, where $M = N^T(T + T_E)$ is the maximal number of agents carrying the information originated at time T. After the expansion, we found a dying out phase with duration $T_D = T_L - T_E$. The whole average lifetime $\langle T_L \rangle$ of information in a system depends on the number of vehicles proceeding in the town. See Fig. 5.4 b. The curve has a minimum. In case of small systems, more agents result in lower lifetime T_L due to the increasing number of updates (younger information) related to Fig. 5.2 b. In order to understand the increasing regime, we have to study the maximum of $N^T(t - T)$ denoted by M. Figure 5.4 c illustrates that for large system $\langle M \rangle / N$ increases linearly with N, so an increasing proportion of agents is carrying the given timestamp in the whole population. If there are more smart vehicles in the city due to the fast spreading a few information timestamps dominate. There can be hundreds of different information in the system, however a dominant part of agents carries

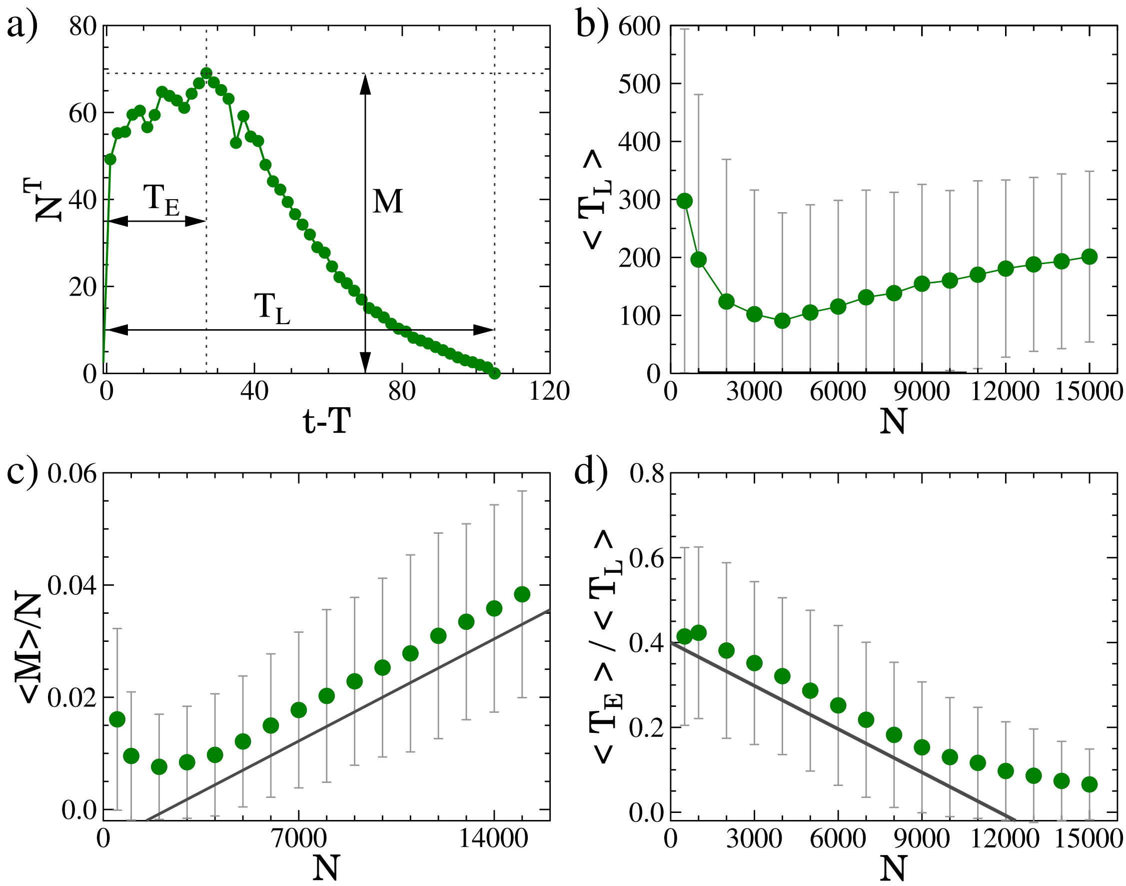

Fig. 5.4. a) The average time evolution of the number of vehicles carrying information with a given timestamp (T) for $N = 5000$. A shorter expansion period T_E is followed by a longer dying out period separated by a maximum M at time $t = T + T_E$. b) The average lifetime of information with a given timestamp as a function of the number of cars. Increasing vehicle density first leads to decreasing lifetime, but above a certain number of vehicles, given information can present in the system for a longer time interval. c) The maximal ratio of vehicles carrying given information with timestamp T at $t = T + T_E$ is increasing linearly for large systems. Solid, gray line just guides the eyes indicating linear dependence. d) The ratio of the expansion period and the lifetime as a function of the number of agents N. The expansion phase is shrinking by increasing the number of agents (indicating speeding up of the spreading since M is increasing).

only a few timestamps. In case of a given timestamp the number of agents carrying it can be a few percent of the population, so hundreds of vehicles. The time needed to overwrite all of them is long, thus the average lifetime of timestamps $\langle T_L \rangle$ can increase with N.

The expansion phase is usually shorter than the dying out phase, but their ratio is not fixed. The $\langle T_E \rangle / \langle T_L \rangle$ ratio as a function of the N number of smart vehicles can be characterized by an almost linear decreasing curve in the studied systems. (See Fig. 5.4 d.) In case of $N = 10^3$ agents, more than 40% of the lifetime of the average timestamp is in the expansion phase, while in case of $N = 10^4$ only 13% of the lifetime is spent in the expansion period. A crowded traffic system results in a short and really intensive spread of new information and it is followed by a long disappearing section with the presence of a few vehicles having old (maybe no longer valid) information.

6. SUMMARY

In this work, we presented an agent-based model of information spreading on Vehicular Ad hoc NETworks. The time-dependent network topology of agents was based on the motion of communicating smart vehicles. Vehicles are moving based on shortest travel time paths between the randomly selected starting and destination points of a real city map. Due to the short-range communication moving vehicles can receive public information from each other or from a fixed Road Side Unit. In this ad hoc network, the statistical properties of information spreading were investigated. Above a certain number of smart vehicles in the system information spreads very fast, and a dominant part of the system can be in an almost homogeneous informed state. This efficient dissemination of information does not require considerable computational power of devices because there is no addressing or routing process and the devices store only the latest packages. However, we found that some agents can stay in an out-of-date state for a quite long time. The number of smart vehicles has a huge effect on spreading, only a large self-organized system can be effective.

On micro scale, we analyzed the time intervals between information exchanges for all agents. We found that the distribution of these intervals has a power-law form where the density of vehicles has no effect on the exponents. The evolution of the spreading of each information holding any given timestamps was also analyzed. We found out that the lifetime of information can be separated into two periods: an expansion period and a dying out period. We found that the number of agents in these complex systems may affect the information spreading and the ratio of these two periods in significant and well describable ways. All these results claim appropriate treating of out-of-date vehicles. The presented results have shown that the system has many interesting features, although real-life applications require more realistic simulations.

In our further research, we try to find answers to essential, practical questions. What happens if the RSU is removed (turned off) or more than one such units are placed to the system? How to avoid the presence of old (out-of-date, fake) information? What is the effect of the introduction of a Susceptible-Infected-Susceptible (SIS) model [23] (forgetting old information)? How to optimize spreading reducing the number of information exchanges (for energy efficiency), but keeping the system in an up-to-date phase? What is the topology of this ad hoc communication network?

ACKNOWLEDGEMENTS

The publication is supported by the EFOP-3.6.1-16-2016-00022 project. The project is co-financed by the European Union and the European Social Fund.

This work was supported by the construction EFOP-3.6.3-VEKOP-16-2017-00002. The project was supported by the European Union, co-financed by the European Social Fund.

Map data copyrighted OpenStreetMap contributors and available from *https://www.openstreetmap.org*.

REFERENCES

1. Porta, S., Crucitti, P., Latora, V. (2006). The network analysis of urban streets: A dual approach, *Physica A*, 369, pp. 853–866.
2. Jiang, B. (2007). A Topological Pattern of Urban Street Networks: Universality and Peculiarity, *Physica A*, 384, pp. 647–655.
3. Yan, Y., Zhang, S., Tang, J., Wang X. (2017). Understanding characteristics in multivariate traffic flow time series from complex network structure, *Physica A* 477, pp. 149–160.

4. Zeadally, S., Hunt, R., Chen, Y. S., Irwin, A., Hassan, A. (2012). Vehicular Ad Hoc Networks (VANETS): Status, Results, and Challenges, *Telecommunication Systems*, 50, pp. 217–241.
5. Fiore, M., Härri, J., Filali, F., Bonnet, C. (2007). Vehicular Mobility Simulation for VANETs, *40th Annual Simulation Symposium (ANSS '07)*, Norfolk, pp. 301–309.
6. Bátfai, N., Besenczi, R., Mamenyák, A., Ispány, M. (2015). Traffic simulation based on the Robocar World Championship initiative, *Infocommunications Journal*, 7, pp. 50–58.
7. Salvo, P., Felice, de M., Baiocchi, A., Cuomo, F., Rubin, I. (2013). Timer-based distributed dissemination protocols for VANETs and their interaction with MAC layer, *IEEE 77th Vehicular Technology Conference*, Dresden, pp. 1–6.
8. Malla, A. M., Sahu, R. K. (2013). A Review on Vehicle to Vehicle Communication Protocols in VANETs, *International Journal of Advanced Research in Computer Science and Software Engineering*, 3, pp. 409–414.
9. Xu, Q., Sengupta, R., Mak, T., Ko, J. (2004). Vehicle-to-Vehicle Safety Messaging in DSRC, *Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks*, pp. 19–28.
10. Nishtha, D. M. (2016). Vehicular ad hoc networks (VANET), *International Journal of Advanced Research in Electronics and Communication Engineering*, 5, pp. 1003–1008.
11. Gong, J., Xu, C. Z., Holle, J. (2007). Predictive Directional Greedy Routing in Vehicular Ad hoc Networks, *27th International Conference on Distributed Computing Systems Workshops (ICDCSW '07)*, Toronto, pp. 2–2.
12. Ramakrishna, M. (2012). DBR: Distance Based Routing Protocol for VANETs, *International Journal of Information and Electronics Engineering*, 2, pp. 228–232.
13. Sanguesa, J. A., Fogue, M., Garrido, P., Martinez, F. J., Cano, J. C., Calafate, C. T. (2016). A Survey and Comparative Study of Broadcast Warning Message Dissemination Schemes for VANETs, *Mobile Information Systems*, 8714142, pp. 1–18.
14. Busetta, P., Bouquet, P., Adami, G., Bonifacio, M., Palmieri, F. (2003). K-Trek: a peer-to-peer approach to distribute knowledge in large environments, *International Workshop on Agents and P2P Computing*, pp. 174–185.
15. Bicocchi, N., Mamei, M., Zambonelli, F. (2012). Self-organizing virtual macro sensors, *ACM Transactions on Autonomous and Adaptive Systems*, 7, pp. 1–28.
16. Varga, I. (2017). Comparison of Network Topologies by Simulation of Advertising, In: Gusikhin, O., Méndez Muñoz V., Firouzi, F., Mønster, D., Chang, C. (eds.) *Proceedings of the 2nd International Conference on Complexity, Future Information Systems and Risk 2017*, Porto, pp. 17–22.
17. Kocsis, G., Varga, I. (2014). Agent Based Simulation of Spreading in Social-Systems of Temporarily Active Actors, In: Was, J., Sirakoulis, G. Ch., Bandini, S. (eds.) Cellular Automata for Research and Industry 2014. LNCS vol. 8751, *Springer*, pp. 330–338.
18. *OpenStreetMap* (2018). [Online]. Available https://www.openstreetmap.org.
19. Varga, I., Némethy, A., Kocsis, G. (2018). Agent-Based Simulation of Information Spreading in VANET, In: Mauri G., El Yacoubi S., Dennunzio A., Nishinari K., Manzoni L. (eds.) *Cellular Automata for Research and Industry* 2018, LNCS vol. 11115, Springer, pp. 166–174.
20. Varga, I. (2020). A complex SIS spreading model in ad hoc networks with reduced communication efforts, *Advances in Complex Systems* (accepted).
21. Stephan Winter (2002). Modeling Costs of Turns in Route Planning, *GeoInformatica*, 6 (4), pp. 345-361.
22. *Google Maps* (2018). [Online]. Available https://maps.google.com.
23. M. E. J. Newman, (2010). *Networks - An introduction*, *Oxford University Press*, pp. 627–675.

