Area-Based Dissemination in Vehicular Networks

Quynh Nguyen
Department of Electrical Engineering
University of Southern California
Los Angeles, California 90036
Email: quynhngu@usc.edu

Bhaskar Krishnamachari
Department of Electrical Engineering
University of Southern California
Los Angeles, California 90036
Email: bkrishna@usc.edu

***Abstract*—Pure opportunistic dissemination of content in a vehicular network can incur high delays if the number of vehicles is relatively low. We consider in this paper an area-based approach to information broadcast in which vehicle to vehicle (V2V) communications is supplemented with vehicle to infrastructure (V2I) communications in order to improve the delay performance. We show how area-based dissemination can analyzed mathematically using a Markovian model. We also investigate through trace-based simulations how different area-partitioning approaches affect the total dissemination time.**

I. Introduction

Vehicular communication has gained an increasing world-wide interest and IEEE 802.11p has been standardized to support Wireless Access for Vehicular Environments. Several consortia (EU C2C-CC [1], US-ITS [2]) have focused on developing key network technologies for communications, security, wireless, routing, etc. for future vehicular communication system. Such vehicular communication systems will enable efficient new safety and infotainment applications for in-vehicle consumption by offloading traffic away from increasingly crowded cellular infrastructure.

Fleet-wide information broadcast for both non-urgent and urgent data is one of the critical building blocks for vehicular networks. For sparse-deployment conditions, a commonly advocated approach has been epidemic propagation using purely vehicle-to-vehicle (V2V) communications [12]. However, precisely due to the sparsity, epidemic propagation can in fact be very slow process. We consider in this paper the possibility of using in addition vehicle to infrastructure (V2I) communication (between vehicles and road-side units) to speed up the dissemination process.

We consider a hybrid V2V/V2I system, in which the infrastructure nodes deployed in a city are organized into distinct sub-networks that we refer to as "areas." We make the assumption that the communications within each area are organized so as to enable very fast propagation of the content to all infrastructure nodes within that area. Vehicles provide the functionality of connecting the areas together, by carrying data between areas.

Figure 1 shows the motivation for this work. This figure is generated based on a real trace of Taxis from Beijing (see section V for more details), and compares the epidemic time of pure V2V information dissemination with the time taken

This work has been supported in part by NSF through award number 1217260

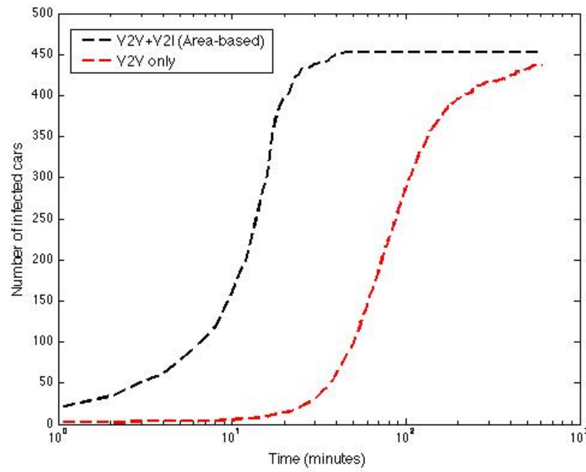

Fig. 1: pure V2V versus area based approach

with V2I-enhanced area-based dissemination in a scenario involving 453 taxis in the central part of Beijing, divided into 36 areas. It can be seen that area-based dissemination of content can be dramatically (order of magnitude) faster.

The information broadcast in this area-based dissemination model works as follows: initially, the content to be distributed is available on a vehicle or infrastructure node within one area (in the former case, the information is first sent from that vehicle to the first infrastructure node it encounters within the area). Once an infrastructure node in that area has the information, it is very rapidly propagated to all infrastructure nodes and vehicles within that area and the information continues to persist within the area. Now, when a vehicle from that area moves to a different area, it communicates its information to other cars and infrastructure nodes it encounters within that area. Again, the information from an infrastructure node in the new area is propagated rapidly to all other infrastructure nodes within that area, and the process continues. In this area-based approach to information dissemination, then, the propagation within areas is very fast, and the latency is primarily dominated by the time taken for vehicles to move from one area that has the content to reach a new area that does not. We are essentially transforming the problem from epidemic propagation between vehicles alone to one of epidemic propagation between areas (with rates determined by the movement of vehicles between them).

We present in this paper a Markovian analytical model for the propagation of content via area-based dissemination. We build on top of a recently proposed model for modeling the mobility of vehicles in a city between areas [8]. To model the information dissemination, we consider as the state of the system a binary vector (of size N, the number of areas) indicating which areas have received the content, and which areas have not. And using the prior work, we derive the transition probabilities between the various states. This forms a transient finite-state continuous Markov Chain with an initial state and an absorbing state. The initial state has a single area, where the content dissemination originates, with "1", and the rest with 0's. The final absorbing state is the vector of all ones, corresponding to all states receiving the disseminated content. Our model allows for a mathematical calculation of the expected time to go from the initial to absorbing state (corresponding to propagating the information to all nodes).

Intuitively, the fewer the number of areas, the quicker the dissemination process will be (since we assume rapid information propagation within the infrastructure for each area). However, in practice, the number of areas will be determined by the cost of establishing the V2I and I2I infrastructure for the sub-networks (initially, each area may even consist of only a single or small number of proximate infrastructure nodes, and slowly expanded, with the number of areas reducing as more infrastructure nodes are deployed over time.) We set aside the issue of how the number of areas should be determined as outside the scope of our investigations, but explore how the propagation time varies with the number of areas and the number of cars. We also examine how the location and structure of the areas affect the speed of information dissemination.

The rest of this paper is organized as follows: section II lists related works; section III describes in detail the problem formulation and its corresponding complexity; section IV illustrates the idea by an example and shows the preliminary results between simulation and theory calculation; section V compares the simulation results of flooding time using different methods of choosing areas; and finally, we present a concluding discussion in section VI.

II. Related Work

Mobility has been considered a crucial aspect which affects accurate connectivity and performance analysis of VANETs [3]. Markov mobility models have been used widely as a strong mathematical tool in evaluating MANETs mobility[4][5]. In VANETs, [6] utilized Markov Chain to predict vehicle directions and velocities, and in a closely-related work, [8] show through validation over real data sets that the Markov Jump mobility model could be used to predict network performance parameters accurately. Further, data dissemination in networks have attracted numerous researchers. An epidemic model based on extracted spatial user mobility of network of devices is developed in [9]. In [14], the authors described a data dissemination scheme by data buffering and retransmitting at intersections of the network. The authors of [10] addressed the fluctuated connectivity of V2V communication by proposing a Voronoi-based placement of RSUs algorithm in V2I networks while in [13], a forwarding algorithm is presented to extend the coverage of Road Side Units in V2I infrastructure network. Our work builds on some of these prior works and makes a complementary novel contribution by analyzing the area-dissemination process taking advantage of both V2I and V2V communications.

III. Problem Formulation

The whole region (say a city) is divided into N areas each of which contains one major intersection. Let the areas be numbered $A_1, A_2, ..., A_N$. There are total M vehicles in the system moving independently among areas in the system.

Vehicles stay in area A_n for an exponentially distributed amount of time with an average time period of $1/\mu_n$ and switch from area A_m to area A_n with probability p_{mn} $(m, n = 1, 2, ..., N)$.

The number of cars in each area, denoted as $W_1, W_2, W_3, ..., W_N$ $(w_i = 1, 2, ..., M)$, follows the steady state distribution $\pi_1, \pi_2, \pi_3, ..., \pi_N$ correspondingly. Let's consider the stable network in which vehicles in the system follow the global balance equation:

$$\sum_{j \neq i} \pi_j (p_{ji} \mu j) = \sum_{j \neq i} \pi_i (p_{ij} \mu i) \tag{1}$$

The probability that there are k vehicles in region i:

$$P(W_i = k) = \binom{M}{k} \pi_i^k (1 - \pi_i)^{M-k} \tag{2}$$

Given that a vehicle has the message initially, the goal is to estimate the epidemic time when the message is disseminated throughout the whole network. Vehicles move around the system from one area to another. We assume that whenever an area gets infected, all infrastructure nodes within the area get the message rapidly and broadcast to all vehicles within the area from that time on (i.e. an "infected" area stays infected). Whenever a vehicle carrying information from an infected area enters a new non-infected area, it sends the piece of information to the nearest road side unit / base station in that area, infecting that area rapidly[1]. This process can be modeled as a Markov Process in which states are the infection status of all areas.

Let say I be the current state. There are currently L_I infected areas in state I, let S_I be the set of infected areas. We already denoted W_i as number of vehicles in area i in the set S_I. We can compute the joint vehicular distribution of all infected areas for state I for a specific case $W = (w_1, w_2, ..., w_{L_I})$ as:

$$\pi_I(W) = P(W_1 = w_1, W_2 = w_2, ..., W_{L_I} = w_{L_I}) \tag{3}$$

$$= \prod_{i=1}^{L_I} P(W_i = w_i) \tag{4}$$

[1]For simplicity, we consider that this time is negligible; however it is straightforward to extend the model to allow for a non-negligible (possibly random) time for intra-area information propagation

Here, w_i can be any value of $0, 1, .., M$. Let us denote all the possible combination set of values for state I as M^{L_I}.

As we consider the transition from state I to J, there will be one more newly infected area, which means obviously $L_J = L_I + 1$. Denote the new infected area as s_J. Denote the area from which the infection happens as s. Because there are L_I infected areas in state I, area s_J can be infected by an infected vehicle coming from any areas in set S_I. The first infected car leaving the infected area s for the non-infected area s_J will spread the infection. Therefore, we can compute the transition probability from state I to state J as:

$$P_{IJ} = \sum_{W \in M^{L_I}} \Bigg(\pi_I(W = (w_1, w_2 ..., w_{L_I})) \tag{5}$$

$$\cdot \sum_{s \in S_I} \Big(\frac{\mu_s \cdot w_s}{\sum\limits_{k \in S_I} (\mu_k \cdot w_k)} P_{s,s_J} \Big) \Bigg) \tag{6}$$

In the above equation, $\sum\limits_{k \in S_I} (\mu_k \cdot w_k)$ is the rate of the first vehicle in the whole region leaving an area for another area, P_{s,s_J} indicates the probability of the vehicle leaving area s for area s_J. Knowing the area distribution of the system, we could compute the mean sojourn time at state I:

$$T_I = \sum_{W \in M^{L_I}} \Bigg(\pi_I(W = (w_1, w_2 ..., w_{L_I})) \cdot \frac{1}{\sum\limits_{k \in S_I} (\mu_k \cdot w_k)} \Bigg) \tag{7}$$

E_I is the expected propagation time starting from state I till it reaches the final state, and let say L is the number of all states in the problem. For example,$L = 8$ in Figure3 in the illustrative example. We can construct a linear equation system of L equations:

$$E_I(1 - P_{II}) = \sum_{\substack{I,J \in L \\ I \neq J}} (T_I + E_J) \cdot P_{IJ} \tag{8}$$

If I_0 is the initial state where exactly one area is infected, the final epidemic time is given as:

$$EP = T_{I_0} + E_{I_0} \tag{9}$$

From the theoretical analysis, the computational complexity of calculating the expressions in equation (5) and equation (7) are $\mathcal{O}(MN^3)$,$\mathcal{O}(MN^2)$ respectively. The complexity for solving the linear equation system (8) is $\mathcal{O}(4^N)$.

IV. Illustrative Example

An illustrative example is given in Figure 2. There are total of 4 areas. Figure 3 shows an example of a corresponding Jump Markov Chain diagram. Infection status of each area in Figure 3 is marked as 1 (all of vehicles in the area have the message) or 0 (there is at least one vehicle in the area which does not have the information). For example, state 2 “1100” implies that area 1 and 2 are already infected, but area 3 and 4 are not.

Figure 4 shows that the theoretical analysis and corresponding simulation results of total propagation time in the illustrative example while total number of vehicles varies from under 50 upto 150 match well with each other. Moreover, figure 5 depicts the close match between theory and simulation for the area distribution of those 4 areas.

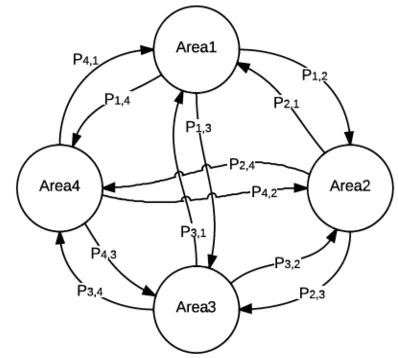

Fig. 2: An example network of 4 areas

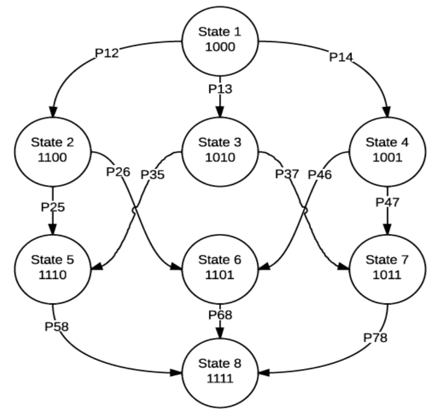

Fig. 3: Corresponding Markov Jump Process

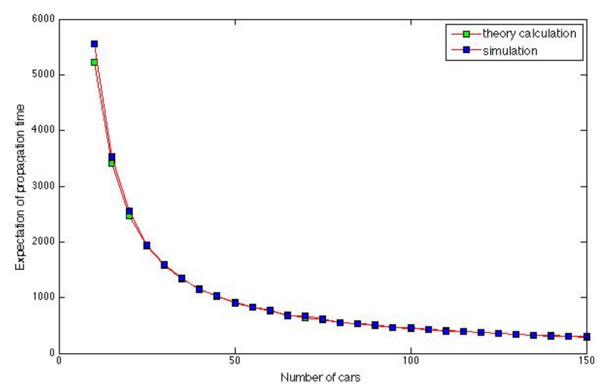

Fig. 4: Expectation of propagation time in theory and simulation for varying number of vehicles

V. Simulation results

A. Simulation methodology

All our simulations and evaluations in this paper are based on a real data set consisting of GPS traces of taxis in Beijing. The data trace contains 272253 GPS points spanning from 39.7082°N to 40.0963°N in latitude, and from 116.2010° to 116.5983°. We extracted critical input parameters for the model (transition probabilty matrix, service) from the data set coordinates of 453 taxis in the high level traffic period from

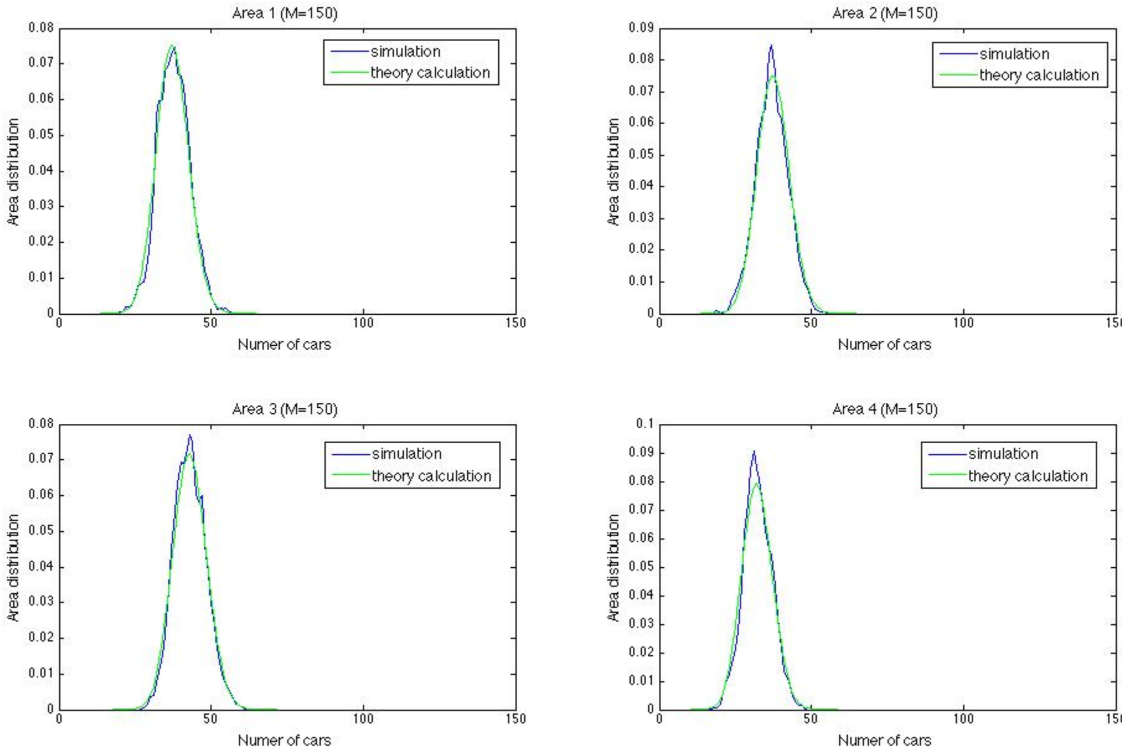

Fig. 5: Area Distribution in theory and simulation in the illustrative example

09:00 am till 07:00 pm recorded approximately every minute in Beijing. The input information is used to convert the model into a Jump Markov Process based on equations 5 and 7 as the input parameters for our simulation.

The number of areas N takes value of 4, 9, 16, 25, 36, 49, 64 areas in the simulations. The number of cars M takes values of 200, 300, 400 and 453; we choose random subsets of the total data set for the smaller numbers of cars.

Area Partitioning Schemes: First, we mark 500 major intersections in the bounded area. The whole bounded region is divided into N areas, each area containing many base stations in multiple major intersections in the same neighborhood cooperating with each others. In the following simulations, we conduct experiments based on 4 different ways of choosing the areas (illustrated in figure 6):

- Grid placement Scheme
- Random Voronoi Scheme: with random choice of intersections from the list [7]
- Density-based Voronoi Scheme: with the choices of intersections in the most busiest areas
- K-means Scheme [11]

For the grid placement scheme, the whole region is divided into a grid of cells with equal edge length. For the next two schemes, given coordinates of the N major intersections, Voronoi partition helps to decide the specific area (Voronoi cell) for each vehicles in the real data trace at specific time. The only difference is that the major intersection is chosen randomly in the random Voronoi scheme but based on densities of vehicles in the latter scheme. For the last scheme, only N is specified. The standard K-means algorithm is used to partition the whole region into N clusters. For a general framework, K-means algorithm can be used as an approach because only N is required.

TABLE I: Mean total propagation time (N=5)

M	Model sim. time	Model cal. time	Empirical time
200 cars	548	555	447
300 cars	341	348	302
400 cars	265	261	254
453 cars	229	227	198

B. Simulation results

Table I compares the mean total propagation time obtained from the model via exact numerical calculations, via model-based simulation (both using fitted parameters from the Beijing data set) with the total propagation time obtained from directly simulating the process on the empirical real data traces. As expected, the model-based calculations and simulations match exactly (providing a further verification of correctness of the analysis). The empirically obtained total propagation times are found to be within 3 to 20% of the model-predictions in all cases.

Figures 7 shows the comparison results among 4 schemes with varying number of divided areas. We can see clearly that following the density-based Voronoi placement scheme obtains the best performance. The total propagation time for utilizing this scheme is the least among all of the 4

(a) Grid placement scheme

(b) Random Voronoi scheme

(c) Density-based Voronoi scheme

(d) K-Means scheme

Fig. 6: Illustration of various area partitioning schemes

presented schemes. Grid placement is much worse compared to all other 3 schemes in which flooding time for a given information piece to reach all of the nodes in the network is much longer on all cases of number of areas due to the fact that intersections in the Beijing area do not match the strict architecture of grid placement, and the vehicle densities area also vary significantly through the grid cells. Moreover, it's also harder to predict the performance of this scheme. The other 2 schemes have the average performance, even though the random-based Voronoi placement seems to have a better performance in some cases.

There is a common trend for all schemes, as anticipated: the less the number of areas, the less the total propagation time needed to spread the information throughout the network. As mentioned before, in practice the number of areas will be determined by extrinsic factors such as the cost of deploying and connecting the infrastructure nodes in the V2I network.

Figures 8 shows the comparison results among 4 schemes with varying number of cars selected from the data set. The more number of cars, the less time required to flood the information piece through the network. The reason is easy to see: the more connectivity among vehicles in the network, the easier it is for one of the cars to move from one area to another to transfer the message.

VI. Conclusions

The deployment of infrastructure in the form of road-side units has the potential to dramatically improve the delay of information broadcast in sparse vehicular networks. We have considered in this paper an area-based dissemination model which utilizes both V2I and V2V communications. Infrastructure nodes in the city are assumed to be organized into autonomous "areas" (sub-networks), such that information propagation to infrastructure nodes and vehicles within each area is fast and persistent, while vehicles store and forward information between areas.

We have presented a tractable Markovian model for area-based information dissemination, and shown that it offers a reasonable approximation to empirical data from a real-world taxi trace. We have used this model to investigate how the placement and shape of areas affects propagation, showing that a density-based Voronoi design yields the best

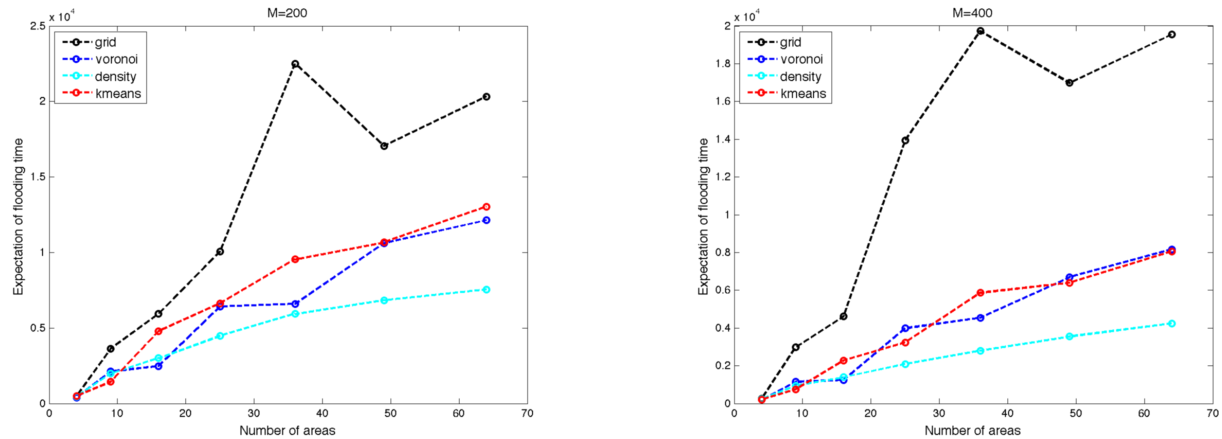

Fig. 7: Propagation time versus number of cars

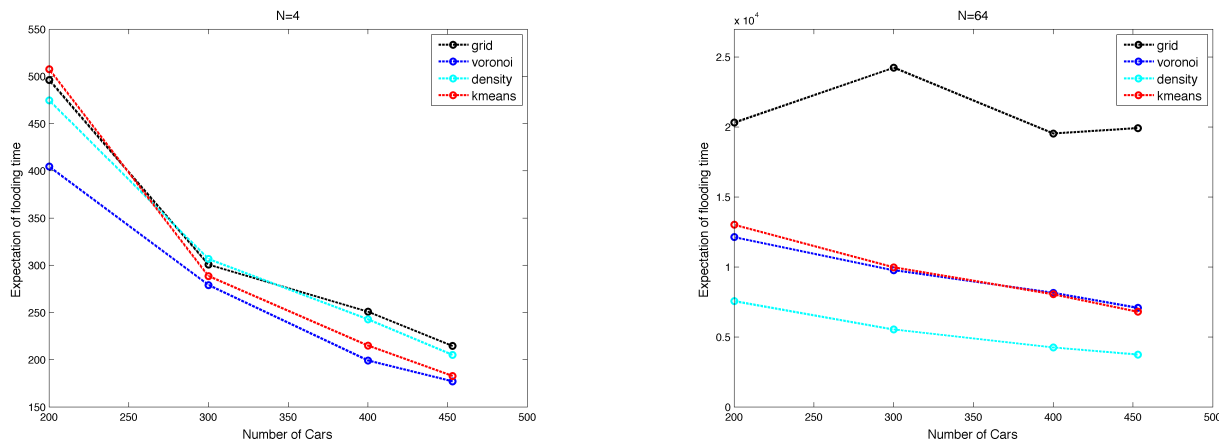

Fig. 8: Propagation time versus number of areas

performance. Further, as expected, the information propagation time is shown to decrease as the number of cars is increased, and as the total number of areas is decreased.

In future work, we plan to dig deeper into the design of the sub-networks within each area, and undertake a more careful accounting of their communication and set-up costs that yields insight into the choice of the optimal number of areas if the proposed approach of areas-based organization of road-side infrastructure were to be followed.

References

[1] Car to car communication Consortium, http://www.car-to-car.org

[2] Intelligent Transportation System, http://www.its.dot.gov

[3] Sarah Madi, Hend Al-Qamzi, "A survey on realistic Mobility Models for Vehicular Ad hoc Networks (VANETs)", *IEEE, Control (ICNSC), 2013, Pages 333-339*

[4] Jinthana Ariyakhajorn, Pattana Wannawilai, and Chanboon Sathitwiriyawong, "A Comparative Study of Random Waypoint and Gauss-Markov Mobility Models in the Performance Evaluation of MANET", *IEEE Communications and Information Technologies, 2006, Pages: 894-899*

[5] Demin Li,Jiacun Wang,Jie Zhou,Peng Yu, "Linking Generation Rate based on Gauss-Markov Mobility Model for Mobile Ad Hoc Networks", *IEEE Communications and Trusted Computing, 2009,Pages:358-361*

[6] Salim Bitam,Abdelhamid Mellouk, "Markov-History based Modeling for Realistic Mobility of Vehicles in VANETs", *IEEE VTC Spring, 2013, Pages:1-5*

[7] Arlind Nocajgoogle maps , Ulrik Brandes, "Computing Voronoi Treemaps: Faster, Simpler, and Resolution-independent", *Computer Graphics Forum, vol. 31, no. 3, June 2012, pp. 855-864*

[8] Yong Li, Depeng Jin, Zhaocheng Wang, Pan Hui, Lieguang Zeng, Sheng Chen, "A Markov Jump Process Model for Urban Vehicular Mobility: Modeling and Applications", *IEEE Transactions on Mobile Computing, 17 Dec. 2013*

[9] Rachit Agarwal,Vincent Gauthier, Monique Becker, "Large Scale Model for Information Dissemination with Device to Device Communication using Call Details Records", arXiv:1305.5675

[10] Prithviraj Patil, Aniruddha Gokhale, "Voronoi-based Placement of Road-side Units to Improve Dynamic Resource Management in Vehicular Ad Hoc Networks", *IEEE Collaboration Technologies and Systems (CTS), 2013 International Conference on , vol., no., pp.389,396, 20-24 May 2013*

[11] http://www.mathworks.com/help/stats/kmeans.html

[12] Fei Ye, Sumit Roy, Haobing Wang, "Efficient Inter-Vehicle Data Dissemination",*Selected Areas in Communications, IEEE Journal on , vol.30, no.4, pp.769,779, May 2012*

[13] Salvo P, Cuomo F.,Baiocchi A.,Bragagnini A, "Road Side Unit coverage extension for data dissemination in VANETs", *Wireless On-demand Network Systems and Services (WONS), 2012 9th Annual Conference on , vol., no., pp.47,50, 9-11 Jan. 2012*

[14] Jing Zhao, Yang Zhang,Guohong Cao, "Data Pouring and Buffering on the Road:A New Data Dissemination Paradigm for Vehicular Ad Hoc Networks", *IEEE Transactions on Vehicular Technology, November 2007*