Quantifying Caching Effects in Urban VANETs

Chaoyi Bian*†‡, Tong Zhao*, Xiaoming Li*, Xiaojiang Du§, Mohsen Guizani¶, and Wei Yan*

*School of EECS, Peking University, Beijing, China
Email: {bcy, zhaotong, lxm, w}@pku.edu.cn
†Venustech, Beijing, China
‡Beijing University of Posts and Telecommunications, Beijing, China
§Department of Computer and Information Sciences, Temple University, USA
Email: dxj@ieee.org
¶Qatar University, Doha, QATAR
Email: mguizani@ieee.org

***Abstract*—Most applications in urban Vehicular Ad hoc NETworks (VANETs) rely on information sharing, such as real-time traffic information queries, advertisements, etc. However, existing data dissemination techniques cannot guarantee satisfactory performance when a lot of information requests come from all around the network. Because these pieces of information are useful for multiple users located in various positions, it is beneficial to spread the cached copies around. Existing work proposed caching mechanisms and conducted simulations for validation, but there is a lack of theoretical analysis on the explicit caching effects. In this paper, we present the cache coverage ratio as the metric to quantify the caching effects, and theoretical analysis is given based on reasonable assumptions for urban VANETs, through which we find the affecting factors include vehicle density, transmission range, ratio of caching vehicles, etc. We deduce the quantitative relationship among them, which have similar forms as the cumulative density function of an exponential distribution. We conduct intensive simulations, which verify the theoretical analysis results match quite well with the simulated reality under different scenarios.**

***Index Terms*—VANET, cache, analysis, simulation**

I. Introduction

Vehicular Ad hoc NETworks (VANETs) are special wireless ad hoc networks in which communication nodes are moving vehicles. VANETs have wide applications in driving safety, intelligent transportation, point-of-interest queries, etc. The key function of these applications is information exchange. For instance, applications of location-related information queries deal with requests such as "*Is there a traffic congestion on XXX Road?*", "*Are there available parking slots near XXX Restaurant?*", etc., which contain the delivery of requests to the corresponding area and responses back.

However, existing routing and data dissemination mechanisms in VANETs cannot support these applications well especially when amounts of requests for the same information come from various positions. Specifically, due to the complex urban environment, packet delivery between distant vehicles is significantly affected by dynamic network topology, intermittent connectivity and features of wireless communication. Existing researches show that the packet delivery ratio presents obvious drops and the latency experiences significant increase when communication vehicles are distant [1][2].

Therefore, caching is beneficial for information sharing by distributing the information or data around the network so that requests from anywhere can be quickly responded by nearby copies. Compared with deploying roadside infrastructure, choosing vehicles as cache carriers has the advantages of low cost and simple configuration. Besides, the routing and data dissemination mechanisms are scarcely affected.

Empirical caching mechanisms are proposed for query applications. The simulation results show that caching brings nearly 100% improvement of the query success ratio in certain scenarios, but different mechanisms have 50% difference in the query delay [3][4]. On the other side, caching introduces overhead on storage and communication for cache management, which leads to a tradeoff between overhead and benefits.

It is necessary to quantify the caching effects on the VANET performance, which not only helps design and assess the caching mechanisms, but also gives a guideline to deal with the overhead tradeoff. However, to the best of our knowledge, there is a lack of related work on studying the caching effects in VANETs with theoretical analysis.

In this work, we try to characterize the caching effects theoretically. We put forward the cache coverage ratio to quantify the caching effect. In brief, the cache coverage ratio is the ratio of vehicles that can access the data within certain hops, while the formal definition will be given later in Sec. II. The main contents of our work include the following.

i) We present the cache coverage ratio to measure the caching effects. Then we give theoretical analysis based on reasonable assumptions of urban VANETs.
ii) We find the quantitative relation among the cache coverage ratio, vehicle density, transmission range and caching rate, which has similar forms as the cumulative density function of an exponential distribution.
iii) We verify the theoretical analysis using intensive simulations, which show that our analysis results match quite well with the simulated reality.

The rest of this paper is organized as follows. Sec. II gives the formal description of the caching problem in urban VANETs. Sec. III shows the theoretical analysis, and Sec. IV demonstrates the simulations. Related work is given in Sec. V. Finally, Sec. VI concludes the paper.

978-1-4673-9814-5/16/$31.00 ©2016 IEEE

II. Problem Statement

In this section, we present the formal description of the caching problem in urban VANETs.

A. Network Model and Assumptions

As described above, VANETs are special wireless ad hoc networks in which vehicles moving on roads are communication nodes. Vehicles communicate with each other through wireless multi-hop transmissions. We apply classic graph modelling for VANETs, i.e., the network is abstracted to a graph $G(V, E)$. V represents the node set, i.e., the set of vehicles, and E represents the edge set. An edge exists between a pair of nodes if the corresponding vehicles can communicate directly. Due to the dynamic network topology, a VANET is hard to be represented by a single static graph. Instead, the network snapshot at any instant can be abstracted to a graph. We try to quantify the caching effects through theoretical analysis on these snapshot graphs which can reflect the communication performance at any instant.

To simplify the modelling, we make some abstractions and assumptions as listed below.

1) We abstract the urban road topology to a Manhattan-like grid, i.e., all roads have the same length, and there are only horizontal and vertical roads.
2) We assume the vehicle headway (i.e., distance between vehicles) follows a certain distribution. Sorts of distributions are proposed by existing work, including exponential distribution [5], Gamma distribution, log-normal distribution [6], etc. Our analysis can be achieved with any distribution, but for simplicity, the exponential distribution is used in the following analysis. In addition, we assume the vehicle density near intersections is higher, which indicates the vehicle headway is exponentially distributed with a larger rate parameter for zones near intersections.
3) We assume the transmission range of a vehicle is a round disk, and all vehicles have the same transmission radius. Vehicles can communicate directly if the distance between them is not greater than the transmission radius. Then, we assume the transmission radius is much larger than the road width so that the affects of road width can be ignored.
4) We assume that each vehicle in the caching area has the same probability to have a cached copy for a specific piece of data, so that caching nodes are uniformly distributed. This probability is equal to the expected caching rate.
5) We do not consider the packet loss due to channel fading or interference in the model analysis.

B. Formulation and Notations

Under the network model and assumptions described above, we propose the cache coverage ratio as follows:

Definition 1. A vehicle is called to be n-hop covered if there exists at least one path of n hops or less between the vehicle and any vehicle with cached data.

In the above definition, the cached data denotes the specific piece of data that we are concerned with.

Definition 2. The n-hop cache coverage ratio of an area is defined as the ratio of vehicles that are n-hop covered, i.e., the number of vehicles that are n-hop covered divided by the total number of vehicles.

The n-hop cache coverage ratio measures the caching effects well, because a high ratio means that most vehicles can get the desired data with no more than n hops, and a low ratio means that amounts of vehicles have to exchange information through a long path.

Then we present the formal problem description of caching effects. In a graph $G(V, E)$ abstracted from a VANET at some instant, nodes are positioned with exponentially distributed intervals on an L-length grid, and an edge exists between two nodes if their distance is not greater than R. The rate parameters for intersection and non-intersection zones are λ_1 and λ_2, respectively. Here we define the intersection zones to be zones within the distance r of intersection points. Besides, each node in the caching area has the same probability μ to be a caching node. The problem is to compute the n-hop cache coverage ratio Φ_n of the caching area. The related notations are listed in Table I for clarity.

TABLE I
NOTATIONS

Notation	Description
L	Grid length of the road network.
R	Transmission radius of nodes.
r	Radius of intersection zones.
λ_1	Rate parameter for intersection zones.
λ_2	Rate parameter for non-intersection zones.
μ	Probability that a node is a caching node.
Φ_n	n-hop cache coverage ratio.

III. Theoretical Analysis

In this section, we present the theoretical analysis to quantify the cache coverage ratio with details.

A. One-hop Cache Coverage Ratio

As a start, we quantify the one-hop cache coverage ratio, which can be computed as shown below.

$$\begin{aligned}\Phi_1 &= |V_C|/|V| \approx E(|V_C|)/E(|V|) \\ &= \frac{\sum_{RS} E(|V_C^{RS}|)}{\sum_{RS} E(|V^{RS}|)} \\ &= \frac{\sum_{RS} P_1^{RS} E(|V^{RS}|)}{\sum_{RS} E(|V^{RS}|)}.\end{aligned} \quad (1)$$

V_C denotes the set of one-hop covered nodes, V^{RS} and V_C^{RS} denote the set of nodes and one-hop covered nodes on a specific road segment (RS), and P_1^{RS} denotes the probability that a node on the road segment is one-hop covered. Since $E(|V^{RS}|)$ can be easily computed based on the exponential distribution, the main concern becomes the computation of P_1^{RS}. A node is one-hop covered if it is a caching node or it is a neighbor of a caching node. Therefore, we have Eq. 2, in

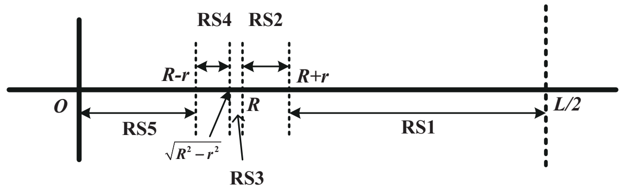

Fig. 1. Division of road segments for computation of cache coverage ratio.

which Q_1^{RS} denotes the probability that a node is a neighbor of a caching node on the road segment.

$$P_1^{RS} = \mu + (1-\mu)Q_1^{RS}. \tag{2}$$

To compute Q_1^{RS}, we have the following theorem.

Theorem 1. *The probability that a node located at x on the corresponding road segment in Fig. 1 has a neighbor caching node can be computed as follows.*

$$Q_1^{RS1}(x) = 1 - e^{-\lambda_2 \cdot 2R \cdot \mu},$$
$$Q_1^{RS2}(x) = 1 - e^{[-\lambda_1(r+R-x)-\lambda_2(x+R-r)]\cdot\mu},$$
$$Q_1^{RS3}(x) = 1 - e^{\left[-\lambda_1(r+R-x)-\lambda_2(x+R-r)-\lambda_1\cdot 2\sqrt{R^2-x^2}\right]\cdot\mu},$$
$$Q_1^{RS4}(x) = 1 - e^{\left[-\lambda_1(3r+R-x)-\lambda_2(x+R+2\sqrt{R^2-x^2}-3r)\right]\cdot\mu},$$
$$Q_1^{RS5}(x) = 1 - e^{\left[-\lambda_1\cdot 4r-\lambda_2(2R+2\sqrt{R^2-x^2}-4r)\right]\cdot\mu}.$$

Due to the space limit, the proof is omitted.

Combining the equations above, the one-hop cache coverage ratio can be computed. Note that for RS2∼RS5, Q_1^{RS} (also P_1^{RS}) depends on the exact node position, thus the expressions of $E(|V_C^{RS}|)$ in Eq. 1 change to integrals as shown below.

$$E(|V_C^{RS2}|) = \lambda_2 \int_R^{R+r} P_1^{RS2}(x)\mathrm{d}x,$$
$$E(|V_C^{RS3}|) = \lambda_2 \int_{\sqrt{R^2-r^2}}^{R} P_1^{RS3}(x)\mathrm{d}x,$$
$$E(|V_C^{RS4}|) = \lambda_2 \int_{R-r}^{\sqrt{R^2-r^2}} P_1^{RS4}(x)\mathrm{d}x,$$
$$E(|V_C^{RS5}|) = \lambda_2 \int_r^{R-r} P_1^{RS5}(x)\mathrm{d}x + \lambda_1 \int_0^r P_1^{RS5}(x)\mathrm{d}x.$$

B. N-hop Cache Coverage Ratio (N>1)

We continue to compute the n-hop cache coverage ratio. Similar to Eq. 1, the ratio can be computed as follows.

$$\Phi_n \approx \frac{\sum_{RS} P_n^{RS} E(|V^{RS}|)}{\sum_{RS} E(|V^{RS}|)}. \tag{3}$$

P_n^{RS} denotes the probability that a node on the road segment is n-hop covered. Also similar to Eq. 2, we have Eq. 4 in which Q_n^{RS} denotes the probability that a node is located in the n-hop range of a caching node on the road segment.

$$P_n^{RS} = \mu + (1-\mu)Q_n^{RS}. \tag{4}$$

However, the determination of the accurate n-hop range is rather complex. Thus we give an upper bound and an approximate computation method instead. The upper bound is based on the fact that the n-hop range with the transmission radius of R cannot be greater than the one-hop range with the transmission radius of nR. As shown below, Φ_n^R denotes the n-hop cache coverage ratio with the transmission radius of R, and Φ_1^{nR} denotes the one-hop cache coverage ratio with the transmission radius of nR.

$$\Phi_n^R \le \Phi_1^{nR}. \tag{5}$$

The approximate method is based on the computation of expected one-hop distance, i.e., the distance to the farthest node within the transmission range. We have the following theorem for the computation of expected one-hop distance.

Theorem 2. *If the transmission radius is R, and the node interval follows the exponential distribution with the rate parameter of λ, the expected one-hop distance $E(D)$ on straight roads can be computed using Eq. 6.*

$$E(D) = R - \frac{1-e^{-\lambda R}}{\lambda}. \tag{6}$$

The proof is also omitted due to the space limit.

Then the n-hop range can be approximated by $2[(n-1)E(D)+R]$. Thus we can compute Q_n^{RS} as follows.

$$\begin{aligned} Q_n^{RS} &= 1 - P(\text{no caching node in the } n\text{-hop range}) \\ &\approx 1 - e^{-\lambda\cdot 2[(n-1)E(D)+R]\mu} \\ &= 1 - e^{-2\mu[\lambda nR-(n-1)(1-e^{-\lambda R})]}. \end{aligned} \tag{7}$$

IV. Simulation

We conduct intensive simulations to verify the theoretical analysis. We choose SUMO [7] (version 0.21) to generate the vehicle traffic, which is based on the realistic microscopic car-following model of vehicles. So the generated mobility trace represents the reality to a certain extent, thus more persuasive in the validation of our analysis. In the following, we will describe the simulation environment and results in detail.

A. Simulation Scenario

We use a Manhattan-grid map with a grid length of 400 m, which contains 30 road segments in each row and also 30 road segments in each column. Thus the total area is 12×12 km^2. The maximum vehicle speed is 40 km/h. Initially vehicles are randomly located in the scenario, and later vehicles only enter or leave the scenario from boundaries. We set a long trip for each vehicle so that the total number of vehicles does not fluctuate greatly during the simulation. Note in this simulation scenario, the vehicle headway is determined by the microscopic car-following model, which is more realistic compared with the exponential-distribution assumption.

B. Validation Results

We present the detailed results below. Each result presented in this subsection is an average of 50 simulation runs.

We first conduct simulations in the scenarios with one lane in each direction and the default SUMO car-following

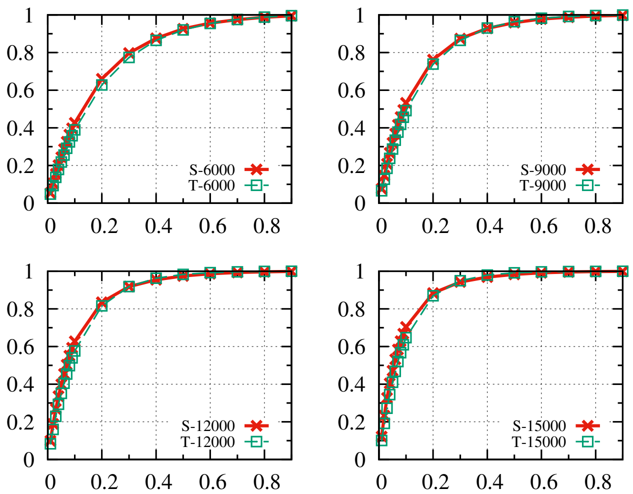

Fig. 2. Comparison between simulations and theoretical analysis with the transmission radius set to 150 m. The x-axis and y-axis represent the caching probability and the one-hop cache coverage ratio respectively. Legend "S-6000" and "T-6000" denote the result for simulation and theoretical analysis with a total vehicle number of 6,000 respectively, similar for others.

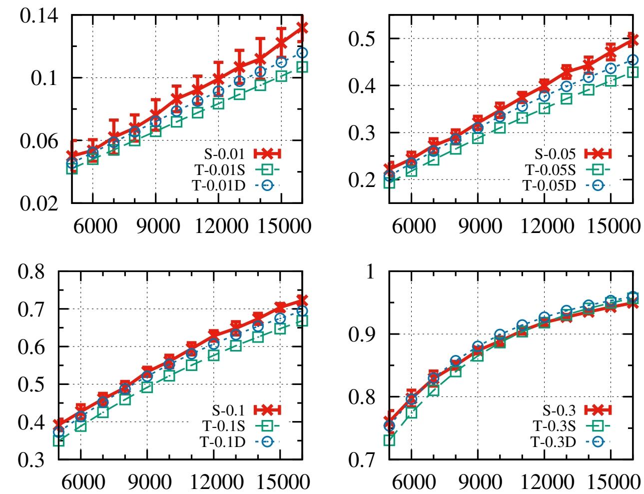

Fig. 3. Comparison between simulations and theoretical analysis with the transmission radius set to 150 m. The x-axis and y-axis represent the total vehicle number and the one-hop cache coverage ratio respectively, and the error bars represent the standard derivation in simulation. Legend "S-0.1" denotes the simulation result with the caching probability of 0.1, while "T-0.1S" and "T-0.1D" denote the theoretical result using the same and different rate parameters respectively, similar for others.

model to verify our analysis. We choose the snapshot at 300 s for analysis because the simulation scenario experiences a variation due to the initial uniform vehicle positioning which becomes steady at about 200 s. We find that there is an obvious difference between the vehicle density around intersections and that in non-intersection zones, and setting the radius of intersection zones to 15 m achieves the greatest difference, which is used in the following results. The rate parameter λ of the exponential distribution used in the analysis is obtained by counting the mean vehicle density in simulations.

Fig. 2 shows the plots for one-hop cache coverage ratios when the transmission radius is 150 m. The x-axis represents the caching probability, and the y-axis represents the cache coverage ratio. As is clearly observed, the results from theoretical analysis match very well with simulations, which verifies that our analysis is in compliance with the simulated reality.

To clearly show the relationship between the cache coverage ratio and the vehicle density, we make plots alternatively in Fig. 3, which also introduces the comparison between using different rate parameters in the intersection and non-intersection zones (legend ended by "D") and the same rate parameter for overall zones (legend ended by "S"). The x-axis represents the total vehicle number, and the y-axis represents the cache coverage ratio. As shown, there is only a slight difference between simulation and theoretical analysis when the caching probability is small. Besides, results of theoretical analysis using different rate parameters always get closer to simulated reality, which proves the distinction for intersection zones is helpful.

The results with other transmission radius are similar, so we omit the plots.

We also verify the two-hop and three-hop cache coverage ratios. Fig. 4 shows the comparison of the simulation results with the theoretical upper bounds and results from the approximate method. As shown, the upper bounds are closer to the simulation results with a larger total vehicle number, because there is higher probability that vehicles exist to forward the transmission so that the n-hop transmission range with the transmission radius of R is more approximate to the one-hop range with the radius of nR. Besides, results from the approximate method match well with simulation, and errors increase with high vehicle density because the approximate method uses the same rate parameters for all zones but intersections become more important for connectivity in dense scenarios.

Besides, we conduct simulations with different lane numbers and other car-following models. The results show that the cache coverage ratio scarcely changes irrespective of lane number and car-following model settings as long as the vehicle density and caching rate remain the same, which means that the validity of our analysis is also not affected.

Then we execute simulations with other street lengths to check whether the street length impacts the validity of our analysis. We keep the entire area of the simulation scenario unchanged, and change the total vehicle number accordingly to maintain suitable vehicle densities. The results are shown in Fig. 5, which prove that the analysis results match well the simulation in the scenarios with different street lengths and vehicle densities. Besides, using different rate parameters in the intersection and non-intersection zones can get much better results compared with using a single one for all zones in scenarios with larger street lengths. Note that the ranges of intersection zones also become larger, e.g., 100 m in 1,200-meter-long street.

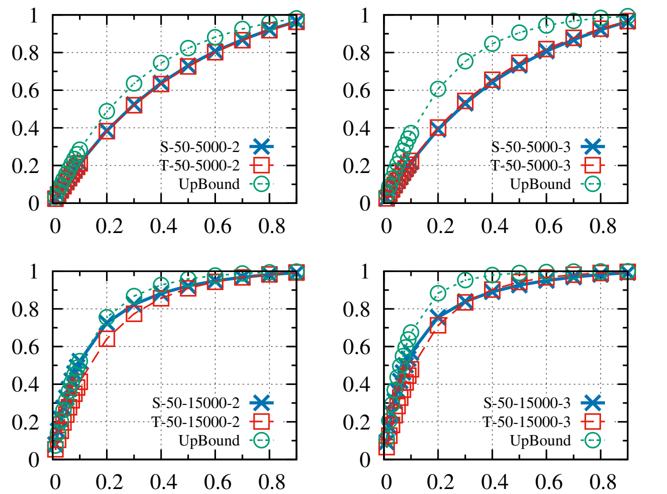

Fig. 4. Results for two-hop cache coverage ratios. The x-axis and y-axis represent the caching probability and the cache coverage ratio respectively. Legend "S-50-5000-2" and "T-50-5000-2" denote the simulation result and approximate computation result for two-hop with the transmission radius of 50 m and a total vehicle number of 5,000 respectively, while "UpBound" denotes the theoretical upper bound, similar for others.

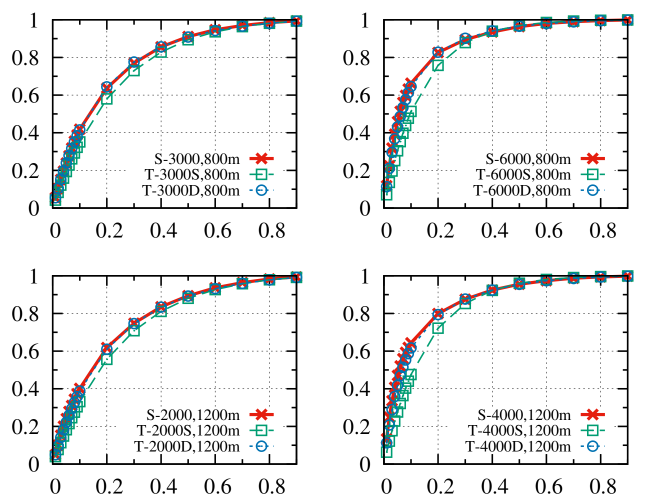

Fig. 5. Results for different street lengths. The x-axis and y-axis represent the caching probability and the one-hop cache coverage ratio respectively. Legend "S-3000,800m" denotes the simulation result with 3,000 vehicles and the street length of 800 m, while "T-3000S,800m" and "T-3000D,800m" denote the theoretical result using the same and different rate parameters respectively, similar for others.

V. Related Work

Some empirical caching mechanisms have been proposed for VANETs. In Hamlet [3], vehicles make caching decisions based on individual observations of data present around. Roadcast [4] considers content popularity for specific applications in cache replacement when the cache buffer is full. Live VANET CDN [8] spreads cached copies around the requester position to amplify the caching effects.

Theoretical analysis for VANETs mainly concentrates on the network connectivity. Existing work [9] introduces percolation theory to study the critical parameters for connectivity, which include vehicle density, transmission radius, etc. Another study utilizes the coverage process to analyze the same problem [10]. There is also work studying the obstacle effects on connectivity [11]. As for caching, only the one-hop mobility influence is discussed [12], which cares about local effects.

To summarize, existing work mainly presents specific caching mechanisms, or conducts theoretical analysis on VANET connectivity, while the analysis for caching effects on the network is missing to the best of our knowledge. Therefore, our work fills this gap as an initial step.

VI. Conclusion

We present the theoretical analysis for caching effects in urban VANETs. The cache coverage ratio is proposed to measure the caching effects in the network, and we find the quantitative relationship among the cache coverage ratio, vehicle density, transmission radius, caching rate, etc. Our analysis results match very well with simulation. Given the vehicle density and transmission range, our analysis model can be used to find a suitable caching rate for desired cache coverage, which is helpful for the design of caching mechanisms. Our future work includes the extension of the analysis to take vehicle mobility into consideration, and the design of caching mechanisms.

Acknowledgment

This work is supported by China NSFC (61272340, 61201245), and 973 Program (2014CB340405), as well as US NSF (CNS-1065444).

References

[1] C. Wu, S. Ohzahata, and T. Kato, "Fuzzy logic based multi-hop broadcast for high-density vehicular ad hoc networks," in *Vehicular Networking Conference (VNC), 2010 IEEE*. IEEE, 2010, pp. 17–24.

[2] C. Wu, S. Ohzahata, and T. Kato, "A broadcast path diversity mechanism for delay sensitive vanet safety applications," in *Vehicular Networking Conference (VNC), 2011 IEEE*. IEEE, 2011, pp. 171–176.

[3] M. Fiore, F. Mininni, C. Casetti, and C.-F. Chiasserini, "To cache or not to cache?" in *INFOCOM 2009, IEEE*. IEEE, 2009, pp. 235–243.

[4] Y. Zhang, J. Zhao, and G. Cao, "Roadcast: a popularity aware content sharing scheme in vanets," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 13, no. 4, pp. 1–14, 2010.

[5] R. P. Roess, E. S. Prassas, and W. R. McShane, *Traffic engineering*. Pearson/Prentice Hall, 2004.

[6] I. Greenberg, "The log normal distribution of headways," *Australian Road Research*, vol. 2, no. 7, 1966.

[7] 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.

[8] N. Nakamura, Y. Niimi, and S. Ishihara, "Live vanet cdn: Adaptive data dissemination scheme for location-dependent data in vanets," in *Vehicular Networking Conference (VNC), 2013 IEEE*. IEEE, 2013, pp. 95–102.

[9] X. Jin, W. Su, and Y. Wei, "Quantitative analysis of the vanet connectivity: theory and application," in *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*. IEEE, 2011, pp. 1–5.

[10] S. Shioda, "Connectivity of vehicular ad hoc networks in downtown scenarios," in *Ad Hoc Networks*. Springer, 2012, pp. 177–192.

[11] M. G. Almiron, O. Goussevskaia, A. A. Loureiro, and J. Rolim, "Connectivity in obstructed wireless networks: from geometry to percolation," in *Proceedings of the fourteenth ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2013, pp. 157–166.

[12] O. Attia and T. ElBatt, "On the role of vehicular mobility in cooperative content caching," in *Wireless Communications and Networking Conference Workshops (WCNCW), 2012 IEEE*. IEEE, 2012, pp. 350–354.