

# Travel Prediction-based Data Forwarding using Realistic Traffic Model in Vehicular Networks

Younghwa Jo  
Department of Software Platform,  
Sungkyunkwan University,  
Republic of Korea  
Email: jyh1127@skku.edu

Youngmi Baek  
Department of Information  
and Communication  
Engineering, DGIST,  
Republic of Korea  
Email: ymbaek@dgist.ac.kr

Jaehoon (Paul) Jeong  
Department of Interaction Science,  
Sungkyunkwan University,  
Republic of Korea  
Email: pauljeong@skku.edu

**Abstract**—Vehicular Ad Hoc Networks (VANET) have become one of the most important research areas for the successful and deployment of intelligent transportation system and the unmanned vehicle technologies. In VANET, Travel Prediction-based Data Forwarding (TPD) was proposed as one of novel data forwarding schemes based on road traffic statistics. However, the TPD could not guarantee good performance in the environment with the traffic light because of the increasing deviation of the road traffic statistics. Since urban road networks have individual traffic lights at each intersection, it is necessary to solve the above problem. In this paper, we devise a method to optimize the TPD in order to make it operate so well in the realistic traffic model. Our performance optimizing method is carried out with a statistical approach and an algorithmic approach. In the statistical approach, we consider a way to control the rise of the deviation in statistics which is the main cause of the significant performance degradation of the TPD. In the algorithmic approach, we first improve the TPD with multi-hop delivery, and next we improve the probability calculating function. Our realistic simulation shows that our statistical and algorithmic optimization method improves the performance of the TPD on a realistic traffic model dramatically.

**Index Terms**—vehicular networks, travel prediction, statistics, optimization, data forwarding.

## I. INTRODUCTION

Vehicular Ad Hoc Networks (VANET) are getting attention as one of promising research areas as they could provide a great potential for efficient, safe and reliable unmanned vehicles in intelligent transport systems. In VANET, the performance of data dissemination highly depends on the traffic mobility pattern, traffic density, and dynamic topology. In order to address this problem, Travel Prediction-based Data Forwarding (TPD) was proposed as one of novel data forwarding schemes [1]–[3]. In TPD, traffic statistics is used to predict encounters of vehicles travelling towards each destination, and a reliable data forwarding path is generated by constructing a graph exploiting these encounters. Therefore, TPD allows the vehicles to efficiently and reliably disseminate various information even under the low density of traffic. Unfortunately, the TPD could not guarantee good performance in the realistic traffic environment with the traffic lights since the TPD has been targeted at the controlled environment without traffic lights. This is because the unpredictable vehicles mobility

related to both traffic lights and various traffic density states by congestion or accidents might highly influence increasing the deviation of the road traffic statistics. It consequently affects its estimate of delivery probability, so that the packet transmission decreases much more than expected.

Especially since urban road networks have individual traffic lights at each intersection, it is necessary to solve the above problems for the adoption of the TPD scheme in order to achieve better performance under real traffic environments in the presence of traffic lights. For this reason we devise a method to optimize the TPD in order to make it operate so well in the realistic traffic model.

Our performance optimizing method is carried out with a statistical approach and an algorithmic approach. In the statistical approach, we consider a way to control the rise of the deviation of expected travel time which is the main cause of a significant performance degradation of TPD. While most of the vehicles pass through the given road network within a certain time span, some vehicles takes more time to travel than the average. Their delays do not happen frequently, but it might have a great influence on the deviation of traffic statistics. Therefore, we introduce a cutting optimization method to generate the statistics with reliability for the TPD. It cuts off the outliers that are so far away from the average and can reduce the deviation in order to increase the statistical prediction accuracy. The algorithmic approach also leads to increase both performance and applicability of the TPD. We can improve in two ways. The first is to provide multi-hop ad hoc routing as an effective way to disseminate packets since the TPD originally supports only one-hop forwarding to a pairwise encounter. Through the multi-hop data forwarding scheme, vehicles can transmit data even if they do not directly communicate with each other. There is therefore no doubt that the delivery probability increases. We introduce Destination-Sequenced Distance-Vector Routing (DSDV) to improve algorithm of the TPD [4]. It is vulnerable to routing error due to vehicular networks with the dynamically and rapidly changing topology since it is one of proactive routing protocols. To address this issue, we add some restrictions on DSDV routing to work well on fast changing VANETs. The second is to expand the range of road segments to be considered for



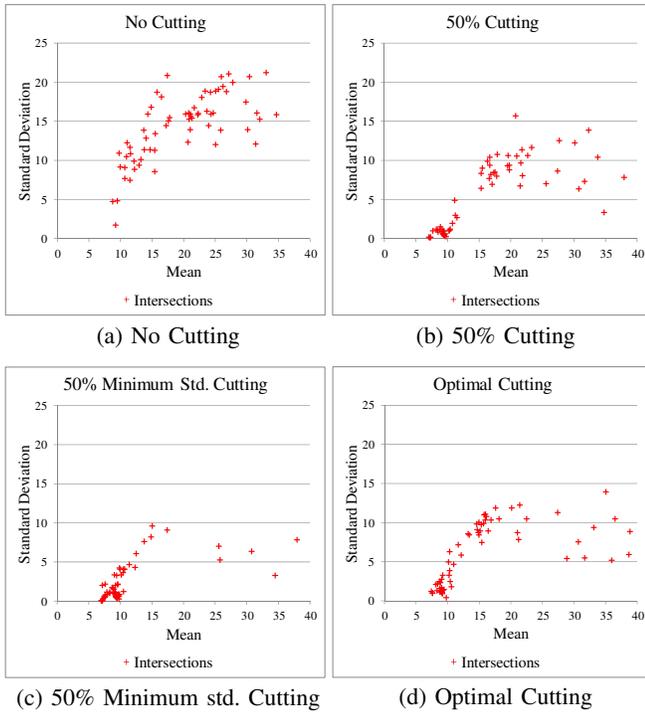


Fig. 2: Mean and Standard Deviation in Road Segments by Statistics Cutting Methods

### A. Statistical Optimization

As mentioned above, the original TPD do not work well in the environment with the traffic lights. This is because it does not consider the vehicle mobility in the presence of traffic lights. For example, its traffic light can stop some vehicles on the road for a certain period of time while vehicles pass through an intersection. At that time, although the travel time of a vehicle at the intersection is only delayed by the amount of time it has been stopped, note that the vehicle travels towards destination under various traffic conditions. Specially, the more traffic density of the intersection increases, the more the delay of the traffic flow dramatically increases. It may make the variance of the travel time increase on the road network. To minimize the variance of total travel time on the real road, we introduce cutting methods for the statistics optimization, so that TPD can perform well on the realistic traffic model.

As shown in Fig. 2(a), which shows the raw traffic data (i.e., the travel time of vehicles travelling each road segment) that are collected by our controlled environment, the standard deviation of measured traffic data almost rises in proportion to the mean value. This rise in standard deviation has a major influence in the actual delivery ratio. A real pairwise encounter between vehicles might not happen as an encounter is predicted by TPD based on vehicle trajectory analysis. Consequently the packet transmission frequently fails, compared with the expected success with the delivery probability derived from traffic statistics in TPD.

To address these serious problems by reducing the standard deviation of the travel time, we try to find out how to

effectively remove invalid data called *outliers* from measured traffic data while maintaining the nature of statistics. If the measured data comes from a symmetric distribution, the outliers have high effect on the mean and the standard deviation. We therefore adopt a trimmed mean method to generate reliable traffic statistics, which is well suited for data with erratic deviation. In order to remove the outliers, the most important thing is to determine an appreciate level trimmed from each tail of the distribution while maintaining statistical reliability.

The first method considered is to remove the largest and smallest values of simply sorted data by an equivalent amount  $\psi$ , where  $\psi$  is the sum of percentage of removed data. This method is borrowed from the method used in sports called *interquartile mean* [15]. In this study, not only the mean value but also the standard deviation is calculated through the truncated part. Data within the interquartile range is generally treated as normal and  $\psi$  in this method is 50%. The result of data trimmed by  $\psi\%$  is shown in Fig. 2(b). We can see that the standard deviation is significantly reduced compared with Fig. 2(a). The second method is similar to the first method, but it finds out the range in which the standard deviation becomes the smallest in order to remove the data out of the range. The standard deviation as shown in Fig. 2(c) has much less value than those of Fig. 2(a) and (b). Unfortunately, there are critical issues to be addressed in the previous two methods. If the TPD is performed based on these statistics, the estimate of the packet delivery probability is very high since most of the data is driven to the minimum road travel time. However, even if vehicles are stopped by one or two traffic lights, the transmission failure might substantially increase. Therefore, a new method to exclude outliers is devised to optimally select traffic data used for traffic statistics for the TPD, according to the traffic flow on a certain road, rather than determining the static range for data to be trimmed. This optimal cutting method is supposed to select only traffic data with travel time in a given range of  $\mu_i \pm \sigma_i$  on each road segment  $i$  with mean  $\mu_i$  and standard deviation  $\sigma_i$  in order to generate traffic statistics. In other words, the dynamic range for data to be trimmed is determined differently for each road section, in which the sum of percentage of removed data varies by each road section. From the result of the optimal cutting with  $\psi_i$  as shown in Fig. 2(d), it turns out that it is better distributes the mean values of travel time on each road section without a wide fluctuation of the standard deviation. Furthermore it shows a similar tendency of the standard deviation with that of Fig. 2(c) and the distribution of the mean is also similar to that of Fig. 2(b).

### B. Algorithmic Optimization

In this subsection, we introduce an algorithmic optimization that enables TPD to achieve better performance in realistic traffic environments, which is consisted of a multi-hop routing method and a multiple road segment method. It results in improving both the expected delivery ratio and actual delivery ratio.

1) *Vehicular Multi-Hop Distance Vector Routing*: originally, the TPD has been implemented using only one-hop transmission, where vehicles can only exchange data whenever they come into the transmission range of each other. In TPD, two vehicles cannot communicate with each other when they are n-hop far away although one vehicle is in close proximity to another. In order to improve connectivity in TPD it is necessary to provide multi-hop transmission capability for these vehicles whose transmission ranges do not overlap each other, so that vehicles can be indirectly connected to someone far away. However, since DSDV is one of proactive routing protocols, it is not well-suited for VANETs in which the topology is highly dynamic and changeable, depending on the velocity, density, trajectory of vehicles on the road. For example, in the VANETs, even if a vehicle is selected as the next hop towards destination from the routing table, it can disappear within a few seconds. To address structural vulnerability of DSDV, we add some constraints on the DSDV routing protocol, to control the rapid changing routing state. The first constraint is related to the routing table update interval and timeout. Update interval of their routing table representing vehicles connectivity should be reduced to very small in order to respond to the dynamic nature of VANETs. The information of the routing table is frequently updated and would be constantly up to date. In consideration of the mobility of vehicles, a smaller value of the route timeout is also set to shorten the previous table entries and to provide more reliable routes. In our simulation, we use 300ms for the update interval. The second constraint is not to exceed the maximum number of hops on multi-hop transmission. For the same reason as mentioned above, while transmission based on ad hoc routing is attempted through the long route to destination, the probability of transmission failure can increase.

2) *Encounter Estimation on Multiple Road Segments*: first of all, in order to explain this part, it is necessary to know how to calculate the encounter probability in TPD. When a vehicle  $V_a$  moves along a path to pass from one intersection  $n_1$  to another intersection  $n_2$ , a vehicle  $V_b$  moves in opposite direction of  $V_a$ , which has a reverse path (from the intersection  $n_2$  to the intersection  $n_1$ ) of  $V_a$ . From here, we represent the expected arrival time of vehicle  $V_x$  on intersection  $n_y$  as  $T_{x,y}$ . The probability,  $P(V_a \oplus_{1,2} V_b)$ , that both  $V_a$  and  $V_b$  might meet each other on one road section connecting two intersections  $n_1$  and  $n_2$  is calculated as follows:

$$P(V_a \oplus_{1,2} V_b) = P(T_{a_1} \leq T_{b_1} \cap T_{a_2} \geq T_{b_2}) \quad (1)$$

It can be seen here that the TPD has considered the possible encounter on only one road segment. Therefore, we expand one road segment for encounter to many road segments in order to increase all possible encounters between vehicles. The encounter calculation formula of equation (1) is changed as follows:

$$P(V_a \oplus_{1,k} V_b) = P(T_{a_1} \leq T_{b_1} \cap T_{a_k} \geq T_{b_k}) \quad (2)$$

When the vehicle  $V_a$  moves along a path to pass  $n_1, n_2, \dots$ , and  $n_k$ , and vehicle  $V_b$  moves along a reverse path ( $n_k, n_{k-1}, \dots$ , and  $n_1$ ) of  $V_a$ , the probability that  $V_a$  and  $V_b$  meet each other on one road section for a whole path ( $n_1, n_2, \dots$ , and  $n_k$ ) is calculated by using equation (2).

In this case,  $T_{a_k}$  (or  $T_{b_1}$ ) is the time traveled by the path of the vehicle further from  $T_{a_1}$  (or  $T_{b_k}$ ), we can express  $T_{a_k}$  (or  $T_{b_1}$ ) as  $d_{x,y}$ , which represents the length of the road segment between intersection  $n_x$  and  $n_y$ .

$$T_{a_k} = T_{a_1} + \sum_{i=1}^{k-1} d_{i,i+1} \quad (3)$$

$$T_{b_1} = T_{b_k} + \sum_{i=k}^2 d_{i,i-1} \quad (4)$$

By equation (3) and (4),

$$T_{a_k} = T_{a_1} + t_{1,k} \quad (5)$$

$$T_{b_1} = T_{b_k} + t_{k,1} \quad (6)$$

where  $t_{1,k} = \sum_{i=1}^{k-1} E[d_{i,i+1}]$  and  $t_{k,1} = \sum_{i=k}^2 E[d_{i,i-1}]$ . Here, if we substitute  $T_{a_k}$  and  $T_{b_k}$  in (2) to (5) and (6), we can obtain the following equation:

$$P(V_a \oplus_{1,k} V_b) = P(T_{a_1} \leq T_{b_1} \leq T_{a_1} + t_{1,k} + t_{k,1}) \quad (7)$$

Based on (7), the probability density function (PDF) for  $T_{a_1}$  and  $T_{b_1}$  is obtained as follows:

$$P(V_a \oplus_{1,k} V_b) = \int_0^\infty \int_x^{x+t_{1,k}+t_{k,1}} f(x)g(y)dydx \quad (8)$$

In this case, since  $t_{1,k} + t_{k,1} \geq t_{1,2} + t_{2,1}$  is always established for  $k \geq 2$ , the encounter probability is increased because the integration area for  $f(x)$  is widened.

#### IV. PERFORMANCE EVALUATION

In this section, we explain the simulation environment and implementation for evaluation and the result from the simulation. For the implementation of realistic traffic model, we implemented the TPD on a proven simulator, Veins [5]. The Veins simulator integrates OMNET++, a proven network simulator, and SUMO, a proven road environment simulator, to realize a VANETs environment [6], [7].

We collect statistical information on the road networks by installing a loop detector on each road, and then we implement the TPD by using this statistical information. We also implement the performance optimizing methods as described in Section III.

Since the most important purpose of the TPD is reliable packet delivery, we use the packet delivery ratio for the performance evaluation. Here, we measure the expected delivery ratio and the actual delivery ratio, respectively, to represent the performance of the improved TPD. The expected delivery ratio represents the calculated probability of the packet using the probability calculation function of the TPD. Likewise, the actual delivery ratio represents the rate at which packets are

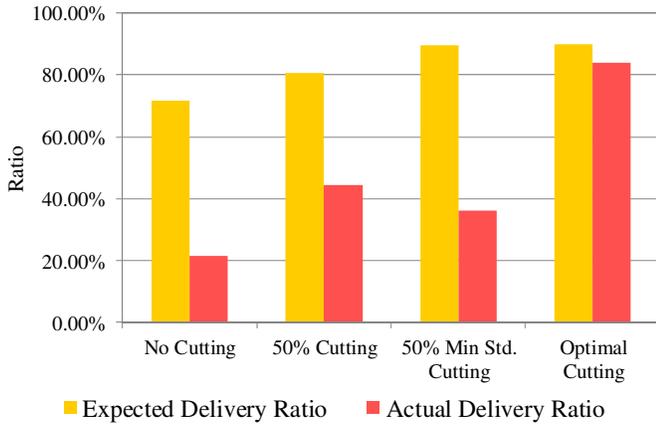


Fig. 3: Expected and Actual Delivery Ratio by Statistics Cutting Methods

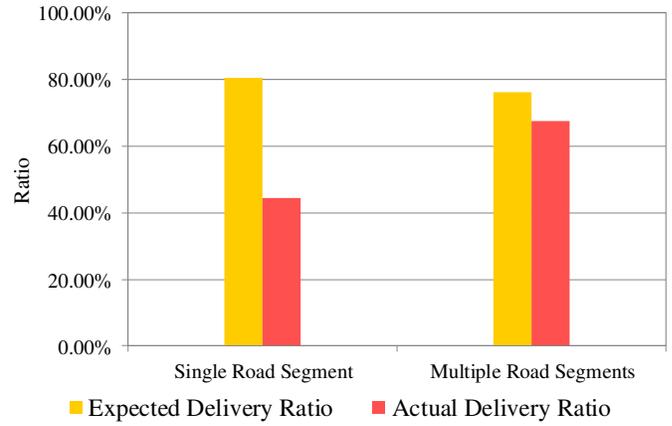


Fig. 5: Expected and Actual Delivery Ratio by Road Segment Methods

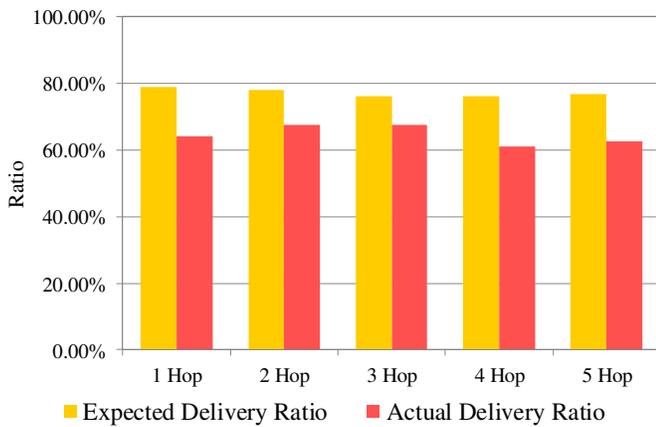


Fig. 4: Expected and Actual Delivery Ratio by Maximum Hop Constraint

successfully transmitted to the destination when the packet is actually transmitted. When evaluating the performance, we should look at these two values together. The higher the actual delivery ratio, the better the performance. However, for reliable delivery, it is best to have a value similar to the expected delivery ratio even if the actual delivery ratio is a little lower. If the actual delivery ratio is higher than the expected delivery ratio, the probability calculating function underestimates the delivery ratio. Similarly, if the actual delivery ratio is lower than the expected delivery ratio, the probability calculating function overestimates the delivery ratio. Therefore, the actual delivery ratio should be close to the expected delivery ratio while maintaining a certain level.

We evaluate our performance optimizing method from three viewpoints. First, we evaluate the statistical optimization method, the statistics cutting method. In this case, we consider four cases: (1) no statistics cutting method (using original statistics, control group), (2) the 50% cutting method, (3) the 50% minimum standard deviation cutting method, and (4) the optimal cutting method.

The evaluation result of statistics cutting method is shown

in Fig. 3. If the cutting method is not used, the expected delivery ratio and the actual delivery ratio are both low. The deviation of the statistics is too high, so it can be seen that the characteristics of the actual traffic cannot be properly applied. The following bar shows the case of using the 50% cutting method. As the deviation of the statistics decreases, the expected delivery ratio and the actual delivery ratio both increase. The third bar shows the case of using the 50% minimum standard deviation cutting method. Here, we can see that the expected delivery ratio increases but the actual delivery ratio decreases. This indicates that overestimation of the probability is made using overly dense statistics to lower the standard deviation. In addition, due to the overestimate, the actual delivery ratio is also lowered by operating differently from the prediction. The last bar shows the optimal cutting method. In this method, the deviation is reduced, but it cuts optimally for each statistics data considering the mean of existing statistics. As a result, the optimal cutting method shows dramatically high performance compared with the results of other methods.

Fig. 4 shows the performance according to the maximum number of routing hops when multi-hop routing is implemented. The one-hop routing means that it does not perform the multi-hop routing because it only allows communication with directly connected vehicles. As the number of maximum hop increases up to three hops, the actual delivery ratio increases. But, over three hops, the actual delivery ratio is decreased, even worse than one hop routing. This is because the vehicles move at a high speed that the routing information changes rapidly. Therefore, if a packet travels through a large number of vehicles in multi-hop routing, the routing may fail in the middle. As a result, we can see from the experimental results that the maximum three hop routing limit is ideal for the TPD operation.

Fig. 5 shows the performance according to the use of multiple road segments method. The single road segment method is a function used in the existing TPD. In this case, the actual delivery ratio is significantly lower than the expected

delivery ratio. This means that the encounter for single road segment cannot calculate the probability accurately. On the other hand, in case of encounter probability for multiple road segment, difference between expected delivery delay and actual delivery delay is small. However, the expected delivery ratio is slightly smaller than in the single road segment.

This is caused by the absolute difference in the number of actually occurring encounters. For the single road segment case, the number of encounters is small, while the encounter probability of each encounter is high. On the other hand, in a multiple road segment case, it is calculated to encounter in a certain probability, even if it is calculated not to encounter in the single road segment case. As a result, the expected encounter probability decreases slightly but the actual encounter probability rises, therefore the performance increases significantly.

## V. CONCLUSION

In this paper, we introduce statistical and algorithmic optimization methods for Travel Prediction-based Data Forwarding (TPD) in realistic traffic model for vehicular networks. Traffic statistics is used to predict encounters of vehicles in TPD. Unfortunately, the TPD could not guarantee good performance in the realistic traffic environment, because of the unpredictable vehicles mobility. Therefore, we devise a method to optimize the TPD in order to make it operate so well in the realistic traffic model. Our performance optimizing method is carried out with a statistical approach and an algorithmic approach. We introduce a cutting optimization method in the statistical approach, and introduce multi-hop routing and expanding to multiple road segments in the algorithmic approach. The evaluation is performed with realistic simulation with Veins simulator, which consists of OMNET++, network simulator, and SUMO, road environment simulator. The simulation shows that our performance optimization method achieves a significant improvement in the performance of the TPD to work well in realistic environment with traffic light.

## ACKNOWLEDGMENT

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1B03035885). This work was supported in part by the Global Research Laboratory Program (2013K1A1A2A02078326) through NRF and the DGIST Research and Development Program (CPS Global Center) funded by the Ministry of Science, ICT & Future Planning. Note that Jaehoon (Paul) Jeong is the corresponding author.

## REFERENCES

- [1] J. P. Jeong, J. Kim, T. Hwang, F. Xu, S. Guo, Y. J. Gu, Q. Cao, M. Liu, and T. He, "TPD: Travel Prediction-based Data Forwarding for Light-Traffic Vehicular Networks," *Computer Networks*, vol. 93, pp. 166–182, 2015.
- [2] J. Jeong, S. Guo, Y. Gu, T. He, and D. H. Du, "Trajectory-Based Data Forwarding for Light-Traffic Vehicular Ad Hoc Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 5, pp. 743–757, May 2011.

- [3] J. Jeong, S. Guo, Y. Gu, T. He, and D. Du, "Trajectory-Based Statistical Forwarding for Multihop Infrastructure-to-Vehicle Data Delivery," *IEEE Transactions on Mobile Computing*, vol. 11, no. 10, Oct. 2012.
- [4] C. E. Perkins and P. Bhagwat, "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers," in *ACM SIGCOMM computer communication review*, vol. 24, no. 4. ACM, 1994, pp. 234–244.
- [5] C. Sommer, R. Geman, and F. Dressler, "Bidirectionally Coupled Network and Road Traffic Simulation for Improved IVC Analysis," *IEEE Transactions on Mobile Computing*, vol. 10, no. 1, pp. 3–15, January 2011.
- [6] A. Varga, "Discrete Event Simulation System," in *Proc. of the European Simulation Multiconference (ESM'2001)*, 2001.
- [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. [Online]. Available: <http://elib.dlr.de/80483/>
- [8] J. Ott and D. Kutscher, "Drive-thru Internet: IEEE 802.11b For "Automobile" Users," in *INFOCOM*. IEEE, Mar. 2004.
- [9] J. Zhao and G. Cao, "VADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 3, pp. 1910–1922, May 2008.
- [10] V. Naumov and T. R. Gross, "Connectivity-Aware Routing (CAR) in Vehicular Ad Hoc Networks," in *INFOCOM*. IEEE, May 2007.
- [11] J. Eriksson, H. Balakrishnan, and S. Madden, "Cabernet: Vehicular Content Delivery Using WiFi," in *MOBICOM*. ACM, Sep. 2008.
- [12] Y. Ding, C. Wang, and L. Xiao, "A Static-Node Assisted Adaptive Routing Protocol in Vehicular Networks," in *VANET*. ACM, Sep. 2007.
- [13] E. M. Royer and C.-K. Toh, "A Review of Current Routing Protocols for Ad-hoc Mobile Wireless Networks," *IEEE Personal Communications*, vol. 6, no. 2, pp. 46–55, Apr. 1999.
- [14] V. N. Soares, J. J. Rodrigues, and F. Farahmand, "GeoSpray: A Geographic Routing Protocol for Vehicular Delay-Tolerant Networks," *Information Fusion*, vol. 15, pp. 102–113, 2014.
- [15] Wikipedia, "Interquartile Mean — Wikipedia, The Free Encyclopedia," 2016, [Online; accessed 28-June-2017]. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Interquartile\\_mean&oldid=731659112](https://en.wikipedia.org/w/index.php?title=Interquartile_mean&oldid=731659112)