Linear Regression-Based Delay-Bounded Routing Protocols for Vehicular Ad Hoc Networks

Yuh-Shyan Chen
Department of CSIE
National Taipei University
Taipei 237, Taiwan, R. O. C.
Email: yschen@mail.ntpu.edu.tw

Chih-Shun Hsu
Department of Info. Management
Shih Hsin University
Taipei, 116 Taiwan, R. O. C.

Yi-Guang Siao
Department of Comm. Eng.
National Taipei University
Taipei 237, Taiwan, R. O. C.

***Abstract*—Routing protocols for vehicular ad hoc networks (VANETs) have attracted a lot of attention recently. Most of the researches emphasize on minimizing the end-to-end delay without paying attention to reducing the usage of radio. This paper focuses on delay-bounded routing, whose goal is to deliver messages to the destination within user-defined delay and minimize the usage of radio. The messages can be delivered to the destination by the hybrid of data muling (carried by the vehicle) and forwarding (transmitted through radio). In the existing protocol, a vehicle may only switch the delivery strategy (muling or forwarding) at an intersection according to the available time of the next block. To improve previous works, our protocol uses linear regression to predict the available time and the travel distance and thus the vehicle can switch to a proper delivery strategy at a proper moment. Therefore, our protocol can reduce the number of relays by radio. Our protocol contains two schemes: the greedy and centralized schemes. The greedy scheme uses only the local vehicle's speed to predict the available time and to decide when to switch the delivery strategy; while the centralized scheme uses the global statistical information to make the decision. Simulation results justify the efficiency of the proposed protocol.**

I. Introduction

Many vehicles are equipped with wireless communication devices nowadays. Based on the idea of mobile ad hoc networks (MANETs), vehicles equipped with wireless communication devices may also form an ad hoc networks named as the Vehicular Ad Hoc Network (VANET). Vehicles on VANETs may communicate through inter-vehicle communication (IVC) or roadside-to-vehicle communication (RVC) [1]. Through IVC and RVC, a vehicle in a VANET would be able to get the information of real-time traffic and emergent notification and thus improves road safety.

Routing protocols for VANETs have attracted a lot of attention recently [2][3][4]. Most of the researches emphasize on minimizing the end-to-end delay. However, different applications have different requirements for end-to-end delay. The notification of car accident and emergency is urgent and needs to be sent to the destination immediately; while file transfer and e-mail can tolerate longer delay time and thus can be considered as lower priority messages.

Since radio is a precious resource, minimizing the end-to-end delay is not that important for lower priority messages. Delivering messages to the destination within the threshold of user-defined delay and minimizing the usage of radio are more important issues for lower priority messages. Therefore, in this paper, we focus on designing an efficient delay-bounded routing protocol, whose goal is to deliver messages to the destination within user-defined delay and minimize the usage of radio so as to save more radio resource for other users.

The delay-bounded routing protocol is first proposed in [5]. Since a vehicle moves much faster than a pedestrian, it assumes that the messages can be delivered to the destination by the hybrid of data muling (carried by the vehicle) and forwarding (transmitted through radio). To minimize the usage of radio, the messages should be carried by the vehicles as long as the time is enough. Two delay-bounded routing schemes have been proposed in [5], the greedy and centralized schemes. In the greedy scheme (named as D-Greedy) messages are delivered along the shortest path; while in the centralized scheme (named as D-MinCost), messages are delivered along the minimum-cost path, where cost stands for the usage of the radio. The greedy scheme has only the average velocity of next block; while the centralized scheme has the average velocity of every block. Therefore, the centralized scheme can apply dynamic programming to calculate the minimum-cost path. However, in both of the schemes, a vehicle may only switch the delivery strategy (muling or forwarding) of messages at an intersection according to the available time of the next block and thus cannot switch to a proper delivery strategy at a proper moment.

To improve previous works, we proposed a novel delay-bounded routing protocol, which uses linear regression to predict the available time and the delivery distance at every sampling moment. Each time after sampling, the predicting line is calculated and the vehicle may switch its delivery strategy according to the predicting line. If the available time is not enough, the delivery strategy can be switched to forwarding, otherwise, the delivery strategy can be switched to muling. This way, a vehicle can switch to a proper delivery strategy at a proper moment. Therefore, our protocol can make a better usage of the available time and reduce the number of relays by radio.

Our protocol also contains two schemes: the greedy and centralized schemes. Both of the schemes are based on liner regression. However, the greedy scheme uses only the local

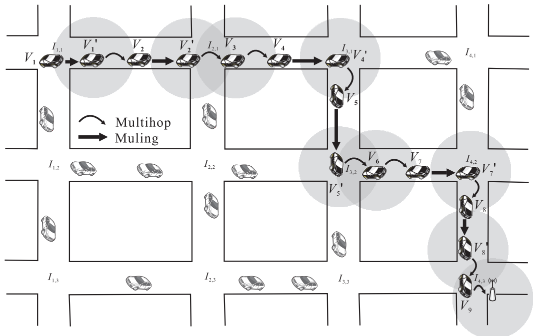

Fig. 1. System model

vehicle's speed to predict the available time and to decide when to switch the delivery strategy; while the centralized scheme uses the global statistical information to make the decision. Simulation results show that our protocol makes a better usage of the available time and performs better than the existing protocol in terms of delivery ratio and the usage of radio.

The rest of this paper is organized as follows. Preliminaries are given in section 2. Section 3 presents the proposed linear regression-based delay-bounded routing protocol. Section 4 evaluates the performance of the proposed protocol. Section 5 concludes this paper.

II. Preliminaries

This section describes the system model and assumptions first, followed by the description of the motivation and basic idea.

A. System Model and Assumptions

The proposed delay-bounded routing protocol is modified from D-Greedy and D-MinCost [5] and is designed for urban area. Du to the limitation of the budget, only a few access points can be deployed in the urban area and these access point cannot cover the whole area. Therefore, a vehicle may need to send messages to the access point via multi-hop communications if it cannot communicate with the access point directly. In our protocol, vehicles are assumed to equip with on-board computers, wireless communication devices, GPS, and digital maps so as to get geographical locations. Access points can only be installed in the intersection and their locations are known by vehicles. Our protocols assume that vehicles can record the travel time and distance in memory. Vehicles can obtain traffic statistic information when contact with access point. When a message is generated by a vehicle, that message is involved with a time-to-live value (TTL). The time-to-live value is considered as a threshold to restrict the message to reach the destination before expired. The goal of our protocol is to make the best usage of the available time and reduce the usage of radio. There are two strategies to deliver messages: data muling (carried by the vehicle) and forwarding (transmitted through radio) as shown in Fig. 1. Switching between the two strategies is a tradeoff between transmission delay and communication cost. Data forwarding increases communication cost but saves delivery time; while data muling increases delivery time but saves communication cost.

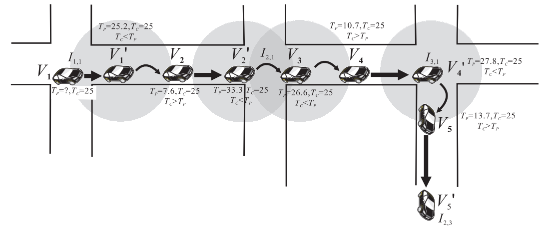

Fig. 2. Routing example of LR-Greedy

B. Motivation and Basic Idea

The major drawback of the existing delay-bounded routing protocol [5] is that it can only switch the delivery strategy at the intersection and thus cannot switch to an appropriate delivery strategy at the appropriate time. For example, if a vehicle determines to forward the message by radio in the block, but the speed of the vehicle becomes high in the middle of the block (or it determines to carry the message by itself, but the speed of the vehicle becomes low in the middle of the block), the vehicle should switch its delivery strategy in the middle of the block. Therefore, our goals are to design a delay-bounded routing protocol which can select an appropriate strategy at the appropriate time and make the best usage of the available time. To achieve our goals, we use linear regression to guide the switch of delivery strategy.

In statistics, linear regression is a regression scheme that models the relationship between a dependent variable Y, independent variables X_i , and a random term ε. The model can be written as the following equation: $Y = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$, where β_i is the respective parameter of the independent variable X_i, and *p* is the number of parameters to be estimated in the linear regression. Since our protocol uses only one independent variable, the linear regression formula is simplified as $\hat{Y} = bX + a$, where $b = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, $a = \bar{y} - b\bar{x}$, $\bar{x}$ is the average of all recorded *x*, and $\bar{y}$ is the average of all recorded *y*.

How we apply linear regression to our protocol is shown as follows:

- Form a criterion line according to the path length D and constrained time T_c.
- Record delivery distance and delivery time periodically.
- Compute the linear regression equation according all the recorded data to form a predicted line periodically.
- If the slope of the predicted line is greater than that of the criterion line, forward the message by radio, otherwise, carry the message by the vehicle.

An example is shown in Fig. 2, where T_c is TTL of the message, T_p is the predicted time calculated by the equation of linear regression. Vehicle V_1 starts at intersection $I_{1,1}$ and uses data muling strategy to deliver message. As the vehicle has arrived $V_1{}'$, T_c becomes smaller than T_p. Vehicle V_1 forwards the message immediately to next vehicle V_2 by radio. The predicted time T_p is decreased because the delivery distance increases greatly in a very short period. So vehicle V_2 can carry the message by itself for a while. When V_2 has arrived $V_2{}'$, T_c is smaller than T_p. The message is forwarded to vehicle V_3 by radio but T_c is still smaller than T_p. So vehicle V_3 forwards the message to vehicle V_4 by radio. Then T_p becomes smaller than T_c and vehicle V_4 carries the message until T_c is smaller than T_p.

III. Linear Regression-Based Delay-Bounded Routing Protocols

We describe our delay-bounded routing protocol in this section. First, we present how to reduce the size of control packets, followed by the description of LR-Greedy scheme and LR-Centralized scheme.

A. Reduce the Size of Control Packets

To make an accurate estimation, the regression line needs to be regenerated periodically according to the latest sampling data. However, as time goes by, the amount of the sampling data will become too large and too costly to be passed to next vehicle. To reduce the size of the control packet, which can be used to generate the regression line, a vehicle needs not to transmit all the sampling data. It only needs to transmit the data which is essential.

As mentioned in section II-B, the formula of linear regression is $\hat{Y} = bX + a$, where $b = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ and $a = \bar{y} - b\bar{x}$. Expand b, we have $b = \frac{(x_1y_1+x_2y_2+\cdots+x_ny_n)-\bar{y}(x_1+x_2+\cdots+x_n)-\bar{x}(y_1+y_2+\cdots+y_n)+n\bar{x}\bar{y}}{(x_1^2+x_2^2+\cdots+x_n^2)-2\bar{x}(x_1+x_2+\cdots+x_n)+n\bar{x}^2}$. We can see that $x_1+x_2+\cdots+x_n$ and $y_1+y_2+\cdots+y_n$ can be derived from $\bar{x}$ and $\bar{y}$, respectively. $\sum_{i=1}^{n+1} x_iy_i$ can be derived from $x_1y_1+x_2y_2+\cdots+x_ny_n$ and the new sampling data (x_{n+1}, y_{n+1}); while $\sum_{i=1}^{n+1} x_i^2$ can be derived from $x_1^2+x_2^2+\cdots+x_n^2$ and the new sampling data x_{n+1}. Therefore, a vehicle only needs to keep (or pass) 5 numbers : $\bar{x}$, $\bar{y}$, n, $\sum_{i=1}^{n} x_iy_i$, and $\sum_{i=1}^{n} x_i^2$. By combining the new sampling data (x_{n+1}, y_{n+1}), a vehicle will be able to generate the new regression line.

B. Greedy Forwarding Scheme (LR-Greedy)

The LR-Greedy scheme uses only the average velocity of the current block and digital map to make prediction. At the beginning, the LR-Greedy scheme uses Dijkstra's algorithm to find the shortest path and then the message is delivered along the shortest path to the destination AP. Assume that the length of the shortest path is D and the TTL threshold is T_c. A criteria line, whose slope is $\frac{T_c}{D}$, can be derived. As the message is delivered by the vehicle either through data muling or data forwarding strategies, the vehicle will record the delivery time (t_i) and delivery distance (d_i) of the message periodically. After each sampling, the vehicle can calculate the regression line according to all the sampling data and then the vehicle will switch the delivery strategy according to the predicted time ($T_p = bD + a$) calculated by the equation of the regression line. As shown in Fig. 3, if $T_p < T_c$, which indicate that the remaining time is enough, the vehicle will carry the message by itself. On the other hand, if $T_p > T_c$, which indicate that the remaining time is not enough, the vehicle will forward the message by radio as shown in Fig. 4.

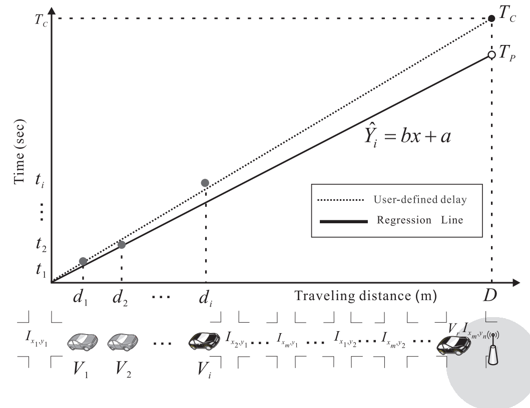

Fig. 3. When $T_p < T_c$, vehicle carries data by itself

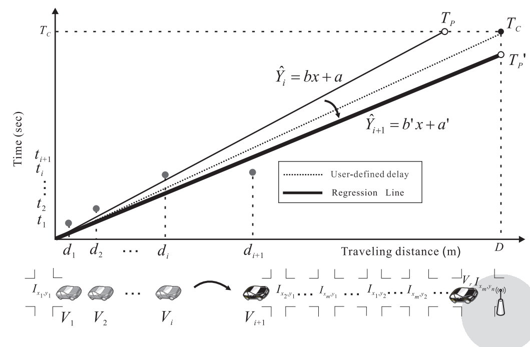

Fig. 4. When $T_p > T_c$, vehicle forward data by radio

C. Centralized Forwarding Scheme (LR-Centralized)

In order to predict the delivery time more accurately and ensure that messages can be transmitted to AP in time, LR-Centralized scheme is proposed. The LR-Centralized scheme assumes that each vehicle has the global information, such as the average velocity and traffic records of each block. With the global information, the vehicle can generate a criterion line for each block and use dynamic programming to find a minimum-cost path, where the cost is defined as the total number of relays by radio.

To find a minimum-cost path, the source vehicle will

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Simulation area	8km $\times$ 8km
Number of a lanes	4
Transmission range	250m
Number of vehicles	200 - 600
Beacon period	5s
Delay threshold	200 - 1100s
Number of generated messages	10
Message size	100KB
Bitrate	1000Kbps
Sampling cycle	0.5s

compute the cost of each block as follows: $cost_{x_iy_i,x_{i+1}y_i} = (BL - V_{x_iy_i,x_{i+1}y_i} \times t_{x_iy_i,x_{i+1}y_i})/T_r$, where BL denotes the block length, T_r denotes the transmission range, $V_{x_iy_i,x_{i+1}y_i}$ denotes the average velocity between intersections I_{x_i,y_i} and I_{x_{i+1},y_i}, $t_{x_iy_i,x_{i+1}y_i}$ denotes the available delivery time from I_{x_i,y_i} to I_{x_{i+1},y_i} in traffic records, $V_{x_iy_i,x_{i+1}y_i} \times t_{x_iy_i,x_{i+1}y_i}$ denotes the possible muling distance of the block, $BL - V_{x_iy_i,x_{i+1}y_i} \times t_{x_iy_i,x_{i+1}y_i}$ denotes the remaining distance which needs to be forwarded by radio.

When the source vehicle is located on intersection I_{x_i,y_j}, the possible next intersections are I_{x_{i+1},y_j} and $I_{x_i,y_{j+1}}$. The recursive function to find the minimum-cost path can be defined as follows:

$$f(x_i, y_j, x_m, y_n) = min \left\{ \begin{array}{l} cost_{x_iy_j,x_{i+1}y_j} + f(x_{i+1}, y_j, x_m, y_n) \\ cost_{x_iy_j,x_iy_{j+1}} + f(x_i, y_{j+1}, x_m, y_n) \end{array} \right. ,$$

$$f(x_m, y_{n-1}, x_m, y_n) = cost_{x_my_{n-1},x_my_n},$$

$$f(x_{m-1}y_n, x_m, y_n) = cost_{x_{m-1}y_n,x_my_n},$$

where the source vehicle is located on intersection I_{x_i,y_j} and the destination AP is located on intersection I_{x_m,y_n}. Dynamic programming can be used to solve the above recursive function and derive the minimum-cost path. The message is then delivered along the minimum-cost path. Similar to LR-Greedy scheme, as the message is delivered by the vehicle, the vehicle also needs to record the delivery time and delivery distance of the message periodically. After each sampling, the vehicle also needs to recalculate the new regression line and compare its slope with that of the criteria line and the result can guide the switch of the delivery strategy. The major difference is that each block has its own criteria line which is derived from traffic records, and thus this scheme can allot a more proper amount of available delivery time to each block.

IV. SIMULATION RESULTS

To evaluate the performance of the proposed LR-Greedy and LR-Centralized schemes, we compare them with D-Greedy and D-MinCost schemes. NCTUns-5.0 [6] is adopted as the simulation tools. The simulation parameters are shown in Table I.

The performance metrics observed in the simulations are:

- *Total transmitted bytes*: the total amount of control and data messages that have been transmitted during the routing process.
- *Delivery ratio*: the total numer of packets that have reached the destination in time divided by the total numer of packets that have been delivered by the source vehicle.
- *Average delivery delay*: the average of the delivery delay of all successful delivered messages within the delay threshold.

The correlation coefficient used in our simulation is defined as a value which indicates the correlation between traffic records and current traffic states. In the simulations, the correlation coefficient is tuned between 0.5 and 1. The greater the correlation coefficient is, the more accurate the traffic records are.

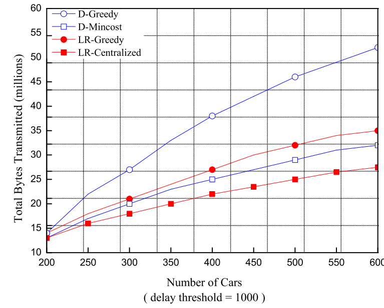

Fig. 5. Total transmitted bytes VS. number of cars

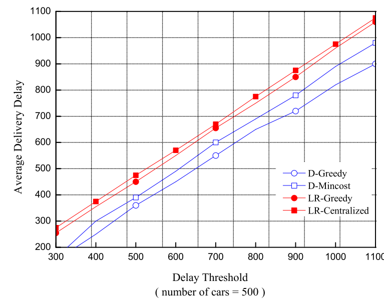

Fig. 6. Average delivery delay VS. delay threshold

A. Total Transmitted Bytes

Fig. 5 shows the impacts of the number of cars to total transmitted bytes. The total transmitted bytes of the LR-Centralized scheme is the the lowest followed by the D-MinCost, LR-Greedy, and D-Greedy schemes. The proposed LR-Centralized and LR-Greedy schemes perform better than the D-MinCost and D-Greedy schemes respectively, because our schemes can switch the delivery strategy whenever the regression line is moving from the upper side to the the lower side of the criteria line (or vice versa). Hence, our schemes can switch to proper delivery strategy at proper moment and thus can reduce the number of relays by radio. As the number of car increases, the total transmitted bytes also increases because higher density of cars will slow down the moving speed of vehicles and thus increases the number of relays by radio.

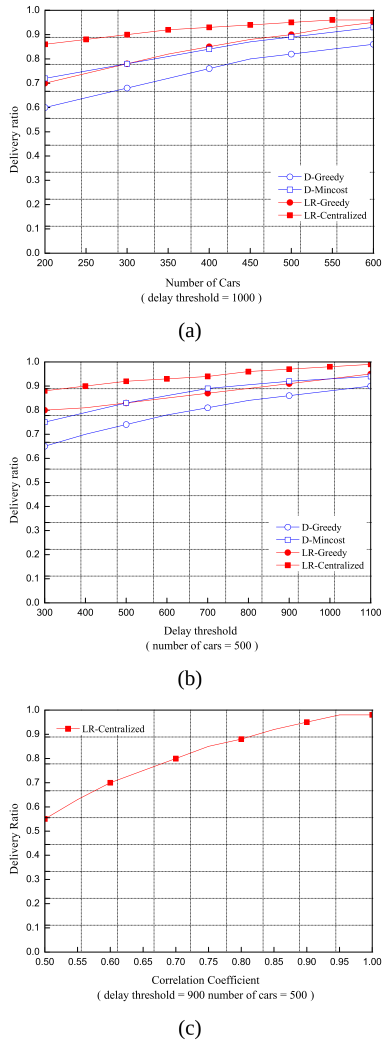

Fig. 7. Delivery ratio VS. (a) number of cars, (b) delay threshold, (c) correlation coefficient

B. Average Delivery Delay

Fig. 6 shows the impacts of the delay threshold to average delivery delay. The average delivery delay of the LR-Centralized scheme is the the highest followed by LR-Greedy, D-MinCost, and D-Greedy schemes. The average delivery delays of the proposed LR-Centralized and LR-Greedy schemes are much closer to the delay threshold than those of the D-MinCost and D-Greedy schemes, which indicates that our schemes can make a better usage of the available time.

C. Delivery Ratio

Fig. 7 shows the impacts of the number of cars, delay threshold, and correlation coefficient to delivery ratio. The delivery ratio of the LR-Centralized scheme is the highest; while the delivery raio of the D-Greedy scheme is the lowest. The LR-Centralized and LR-Greedy schemes perform better than the D-MinCost and D-Greedy schemes respectively, because our schemes can make a better usage of the available time and reduce the frequency of data forwarding and thus improves the delivery ratio. a vehicle can switch to a proper delivery strategy at a proper moment. Therefore, our schemes can make a better usage of the available time and reduce the number of relays by radio and thus achieves higher delivery ratio. As the number of cars increases, the delivery ratio also increases because high density of cars may incur more candidate to relay the message and thus increases the delivery ratio. As the number of cars increases, the delivery ratio also increases because high density of cars may bring more candidates to relay the messages and thus increases the delivery ratio. As the delay threshold increases, the delivery ratio also increases because the vehicle has more available time to deliver the messages. Higher correlation coefficient brings higher delivery ratio because higher correlation coefficient indicates more accurate traffic records and predictions.

V. Conclusions

In this paper, we have presented a delay-bounded routing protocol for vehicular ad hoc networks. Our protocol contains two schemes: the LR-Greedy and LR-Centralized schemes. Both of the schemes use liner regression to predict the available time. However, the LR-Greedy scheme uses only the local vehicle's speed to predict the available time and to decide when to switch the delivery strategy; while the LR-Centralized scheme uses the global statistical information to make the decision. Simulation results show that both of the proposed schemes can make a better usage of the available time and perform better than the D-Greedy and D-MinCost schemes in terms of delivery ratio and the usage of radio. The LR-Centralized scheme possesses more information and thus can make a more accurate prediction. Hence, the LR-Centralized scheme performs the best, but it needs to gather more information; while the LR-Greedy scheme can be applied easily.

References

[1] J. Luo and J.-P. Hubaux, “”A Survey of Inter-Vehicle Communication”,” *EPFL Technical Report IC, Switzerland*, 24 March 2004.

[2] U. L. K. C. Lee, J. Haerri and M. Gerla, “”Enhancde Perimeter Routing Geographic Forwarding Protocols in Urban Vehicular Scenarios”,” *IEEE Global Communications conference , (GLOBECOM 2007), New Orleans, LA, USA, pp. 1-10*, 26-30 November 2006.

[3] K. M. Z. Mo, H. Zhu and N. Pissinou, “”MURU: A Multi-Hop Routing Protocol for Urban Vehicular Ad Hoc Networks”,” *International Conference on Mobile and Ubiquitous Systems , (MobiQuitous 2006), San Jose, California, USA, pp. 1-8*, 17-21 July 2006.

[4] J. Zhao and G. Cao, “”VADD: Vehicle-Assisted Data Delivery in in Vehicular Ad Hoc Networks”,” *IEEE International Conference on Computer Communications, (INFOCOM 2006), Barcelona, Caralunya, Spain, pp. 1-12*, 23-29 April 2006.

[5] A. Skordylis and N. Trigoni, “”Delay Bounded Routing in Vehicular Ad Hoc Networks”,” *ACM International Symposium on Mobile Ad Hoc Networking and Computing , (MOBIHOC 2008), Hong Kong, China, pp. 341-350*, 27-30 May 2008.

[6] S. Wang and C. Chou, “”NCTUns Tool for Wireless Vehicular Communication Network Researches”,” *Simulation Modelling Practice and Theory*, 2009.