Traffic Aware Routing in Urban Vehicular Networks

Ting Cao*, Xinchao Zhang*, Linghe Kong*, Xiao-Yang Liu*, Wei Shu+, and Min-You Wu*

*Dept. of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
+Dept. of Electrical and Computer Engineering, University of New Mexico, Albuquerque, USA

***Abstract*—An urban vehicular network is a typical type of Delay Tolerant Network (DTN). Based on the routing analysis in a DTN, we first put forward a Minimum Delay and Hop Algorithm (MDHA), which requires both historical and future information on all the vehicles in the network. Since MDHA is not practical, we then design a Traffic Aware Routing Algorithm (TARA), which uses the historical and the real-time vehicle information to make routing decisions on the road structure level. A simulation using real GPS data in Shanghai shows that TARA significantly reduces the transmission delay and the hop count compared to the traditional GEO routing and GPSR.**

I. INTRODUCTION

An urban vehicular network is composed of vehicles with short-range wireless communication devices. Since the vehicles are moving all the time, the wireless links between them are established and broken frequently, which results in dramatic changes in the network topology. The sparse vehicle density and the inhomogeneous vehicle distribution also make the network connectivity low. As a consequence, an urban vehicular network shows some characteristics of a Delay Tolerant Network (DTN)[1][2].

Existing DTN routing algorithms can be roughly divided into three types[3]: 1) Flooding-based (broadcast, infectious) routing[4], whose main challenge is to minimize the total number of the data packets during the broadcast. Opportunistic routing[5] derives from flooding and takes advantage of the unreliable wireless links. 2) Mobile-pattern-based routing[6][7], which analyzes the regular encounter patterns between the vehicles from the statistical perspective and inspires probabilistic algorithms. 3) GEO routing[8], which always chooses the geographically nearest node towards the destination as the next hop. In order to avoid local minimum points, the original GEO routing is improved by right hand traversal rules in plane graphs, which derives GPSR[9], VCLCR[10], GPCR[8], GOAFG[11] and GFC[12].

Based on the routing analysis in a DTN, we first put forward a Minimum Delay and Hop Algorithm (MDHA), which takes advantage of global vehicle information to calculate the optimal routing solution. Since MDHA requires both historical and future information on all the vehicles in the network, it is not practical. Nevertheless, the MDHA path can be used as a baseline to indicate the upper bound of the routing performance.

In order to approach this upper bound, we then propose a Traffic Aware Routing Algorithm (TARA), which uses the historical and the real-time traffic information to make routing decisions on the road structure level. The performance of TARA is simulated using a dataset of two-year GPS tracking data from over 4000 taxis in Shanghai. According to the simulation results, TARA gives a significant performance improvement both in the transmission delay and the hop count compared to the traditional GEO routing and GPSR.

The contribution of TARA lies in three dimensions: 1) TARA observes the urban vehicular network from the road structure level instead of the vehicle level, and makes routing decisions from the macro perspective. As a consequence, the network dynamics are reflected in the edge weights instead of the network topology, which makes the network structure more stable. The routing decisions no longer depend on the unpredictable and highly varied behaviors of the individual vehicles, but follow the statistical pattern of the traffic flow upon the road structure instead. 2) TARA uses the historical and the real-time traffic information to predict the tendency of the traffic flow, which enables distributed and intelligent decision making by the individual vehicles. 3) TARA is simulated using a dataset of real GPS tracking data, which makes the simulation more realistic and credible.

The rest of the paper is organized as follows. Section II introduces MDHA. Based on the analysis in Section II, Section III proposes TARA. The simulation results are demonstrated in Section IV. Section V describes some future work.

II. MINIMUM DELAY AND HOP ALGORITHM

If the bandwidth of the vehicle-to-vehicle wireless communication is limited and multiple packet deliveries occur at the same time, it is NP-hard to find out an optimal routing solution. Thus, we assume that the transmission bandwidth is unlimited. With this assumption, we propose MDHA, which seeks the optimal routing solution to minimize the transmission delay and the hop count of one single data packet.

A. Mathematical modeling

This section defines the basic and critical concepts for the following discussions about MDHA.

1) Network representation: As in DTN[3], a weighted directed graph $G(V, E, T, F)$ is used to describe the urban vehicular network:

- *Vertex*: The set of the vehicles in the network is represented by $V = \{v_k | 1 \leq k \leq N_v\}$.
- *Edge*: The set of the wireless links between the vehicles is represented by $E = \{e_{i,j} | 1 \leq i, j \leq N_v, i \neq j\}$.

978-1-4673-5939-9/13/$31.00 ©2013 IEEE

- *Time*: All the timestamps are represented by a finite set $T = \{t_i | t_{i+1} > t_i, 0 \leq i \leq N_t\}$. Usually we choose a fixed interval for discrete time periods, that is $t_{i+1} - t_i \equiv c$, where c is also called the granularity. The total time is then $c \times N_t$ with a fixed granularity.
- *Function*: The edge weight is the transmission cost between the corresponding vehicles, which is usually the transmission delay. Suppose $F : E \times T \rightarrow \Re$ is the mapping function of the edge weights, then

$$\begin{cases} F(e_{i,j}, t) < \infty, \ e_{i,j} \ exists \ at \ time \ t \\ F(e_{i,j}, t) = \infty, \ e_{i,j} \ does \ not \ exist \ at \ time \ t \end{cases}$$

2) Routing and routing path: In general, a routing is to find a time series including several nodes in the network between the source and the destination, so that the data packets can be delivered from the source towards the destination following the time series, which is so called a routing path P.

For example, $P = (\langle v_1, t_0 \rangle, \langle e_{1,2}, t_1 \rangle, \langle v_2, t_2 \rangle)$ means that the node v_1 generates a data packet at the time t_0 and sends it to v_2 through the wireless link $e_{1,2}$ at t_1; and v_2 completes the receiving process at t_2.

3) Routing cost: Similar to the edge weights, the routing path P has its cost $C(P)$. The goal of a routing algorithm is always to minimize some cost of the routing path. If the total transmission delay is chosen as $C(P)$, $C(P)$ is usually larger than the sum of the transmission delays between the vehicles, since packets will also be delayed when residing on a vehicle and waiting for the next transmission.

B. Minimum Delay and Hop Algorithm

MDHA chooses the transmission delay as the edge cost, and tries to minimize the transmission delay and to reduce the hop count at the same time. Since there are multiple objectives to be optimized, MDHA optimizes a target vector instead of a target function:

$$\min_P \overrightarrow{C(P)} = \min_P \left\{ \sum_{(e_{i,j}, t_k) \in P} [F(e_{i,j}, t_k) + (t_k - t_{k-1})], n \right\}$$

where $F(e_{i,j}, t_k)$ is the transmission delay of one hop, $(t_k - t_{k-1})$ is the carrying delay of one node, and n is the number of the transmission hops.

Suppose $MDHA(t_0, v_s, v_d, G)$ is a specific MDHA problem and an optimal solution path is $MDHA_i(t_0, v_s, v_d, G) = (\langle v_s, t_0 \rangle, \ldots, \langle v_k, t_r \rangle, \langle e_{k,l}, t_{r+1} \rangle, \ldots, \langle v_d, t_n \rangle)$.

If the node v_k can receive the packet before the time t_{r+1}, the total transmission delay of $MDHA_i(t_0, v_s, v_d, G)$ will not change. Thus the critical part of MDHA is to adjust t_r to minimize the hop count to reach v_k with the constraint that $t_r < t (= t_{r+1})$.

The shortest path problem in dynamic networks is also known as the time-dependent shortest path problem[13], since the concrete shortest path depends on the starting time of the algorithm.

The Multi-Objective Optimization (MOO) is studied in detail by R. Marler and J. Arora[14]. Since in most cases a global optimization of all the objectives is hardly possible, we constrain the search algorithm by giving delay minimization a higher priority.

Derived from the Time-Space expansion plan[13] and the MOO[14], MDHA is as follows:

Algorithm 1 MDHA

```
$V_a \leftarrow \{v_s\}$ {$V_a$ includes all the nodes that have already received the data packet, $v_s$ is the source node}
$Q \leftarrow \emptyset$ {$Q$ is a temporary set}
$t_i \leftarrow t_s$ {$t_s$ is the starting time point}
for all $v \in V$ do
    $revtime(v) \leftarrow \infty$
    $path(v) \leftarrow \emptyset$
    $path\_till\_now(v) \leftarrow \emptyset$
end for
$revtime(v_s) \leftarrow t_s$
while $t_i \neq t_n$ and $V_a \neq V$ do
    for all $v_i \in V_a$ do
        for all $v_j \in neighbors(v_i)$ do
            if $revtime(v_j) < t_i$ then
                $revtime(v_j) \leftarrow t_i$
                $path(v_j) \leftarrow path\_till\_now(v_i) + v_j$
                $path\_till\_now(v_j) \leftarrow path(v_j)$
            end if
            if $v_j \notin Q$ then
                $Q \leftarrow Q + v_j$
            else
                if $path\_till\_now(v_j).length > path\_till\_now(v_i).length + 1$ then
                    $path\_till\_now(v_j) \leftarrow path(v_i) + v_j$
                end if
            end if
        end for
    end for
    $V_a \leftarrow V_a + Q$
    $Q \leftarrow \emptyset$
    $t_i \leftarrow t_{i+1}$
end while
```

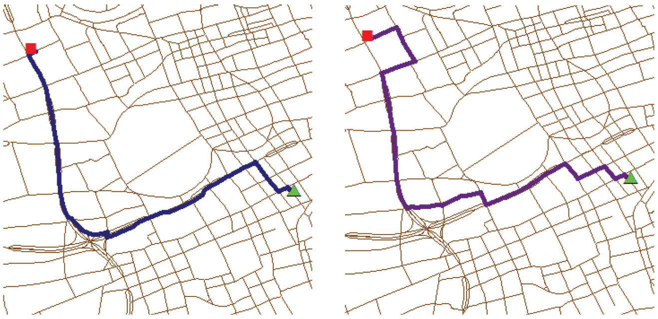
C. MDHA path characteristics

Property 1: Node carrying is a critical behavior during the routing.

Many vehicle-to-vehicle data transmissions occur near road intersections, while the vehicles tend to carry the data packets for a while along the road sections.

The transmission efficiency of almost all the routing algorithms will suffer from a sharp decline in the long-distance scenario. However, *Property 1* indicates that the transmission delay in MDHA is largely sensitive to the following factors: 1) the straight-line distance between the source and the destination; 2) the topology of the road network.

Property 2: The spatial shape of the routing path is more stable than the node series.


(a) Original path (b) Two nodes removed


(c) Four nodes removed (d) Six nodes removed

Fig. 1. Spatial stability of MDHA routing paths

Fig. 1 demonstrates the MDHA routing paths where two, four and six nodes of the original path are removed from the network. The node series of the routing path change rapidly due to the removal of the nodes. For example, the path in Fig. 1(b) contains seven hops, and the path in Fig. 1(c) contains nine hops, while these two paths only share two common nodes. However, the spatial shapes of these paths do not change much.

This property implies, if the source and the destination are determined, the spatial shape of the MDHA routing path will mainly depend on the road structure instead of the concrete vehicle series. Since the behavior of the individual vehicles will not affect the shape of the routing path much, the vehicles can be treated as identical data bearing nodes, which enables a macro routing perspective.

III. TRAFFIC AWARE ROUTING ALGORITHM

MDHA needs both historical and future information on all the vehicles in the network, which cannot be obtained in practice. However, since MDHA produces the optimal routing path, a routing algorithm can achieve better performance if its routing path approximates the MDHA path.

Property 1 shows that the significant vehicle-to-vehicle transmissions of the MDHA path often appear near road intersections. *Property 2* inspires the idea of looking at the vehicular network from a macro perspective. Since the shape of the MDHA path does not change much, it is possible to simulate the MDHA path in practice.

Based on the above knowledge, this section introduces TARA, which finds a stable MDHA routing path and forces data packets to be transmitted along this path.

In TARA, the road intersections instead of the vehicles are treated as the network nodes, and the roads themselves are the network edges. The statistical distribution and the mobility of the vehicles on the road determine the data transmission capacity of the road, and the road capacities are used as the edge weights. The density and the average velocity of the vehicles are changing over time, thus the edge weights are also time-varying, while the road structure remains static. As a consequence, the dynamic details are hidden behind the static structure.

TARA uses the weighted shortest path algorithm to calculate the best routing path in the road network, which is in fact an MDHA routing path with a series of road intersections. The transmission of a data packet is along the calculated intersection series. And geographic routing is used for the data transmissions between two adjacent road intersections. To conclude, the routing path in TARA is no longer a series of vehicles, but a series of road intersections instead.

The critical part of TARA is to determine the edge weights, which are relevant to the traffic flow on the road.

A. Symbols for edge weights calculation

1) r_i*:* the road i

2) L_{r_i}*:* the length of r_i

3) $pd_{r_i}^0$ *and* $pd_{r_i}^1$ *(*Packet Delay*):* the average transmission time along the two directions of r_i

4) $pv_{r_i}^0$ *and* $pv_{r_i}^1$ *(*Packet Velocity*):* the transmission rate

5) $\rho_{r_i}^0$ *and* $\rho_{r_i}^1$*:* the average vehicle density

6) $v_{r_i}^0$ *and* $v_{r_i}^1$*:* the average vehicle velocity

7) $d_{r_i}^0$ *and* $d_{r_i}^1$*:* the average travel time of a vehicle to pass through r_i

If an explicit distinction between the two directions of the road is not required, the digital superscript of the above variables can be omitted, e.g. d_{r_i} generally refers to the travel time along both directions of r_i.

It is obvious that all the variables are functions of time except for L_{r_i}: $d_{r_i} = d_{r_i}(t)$, $pd_{r_i} = pd_{r_i}(t)$, $v_{r_i} = v_{r_i}(t)$, $pv_{r_i} = pv_{r_i}(t)$, and $d_{r_i}(t) = L_{r_i}/v_{r_i}(t)$, $pd_{r_i}(t) = L_{r_i}/pv_{r_i}(t)$.

B. Algorithms for edge weights calculation

We propose three types of edge weighting methods:

1) Length Weighted (LW): The weight of r_i is L_{r_i}. The result path has the shortest travel distance.

Since vehicles carrying packets is the critical behavior in MDHA, the point-to-point routing delay mainly depends on the time for which a vehicle carries a packet. The minimized transmission distance helps minimize the driving time of the vehicles with the data packets, which leads to a reduction of the transmission delays.

2) Vehicle Time Weighted (VTW): The weight of r_i is d_{r_i}, similar to LW. The result path has the shortest travel time.

3) Regression Weighted (RW): The transmission time for a packet to pass through r_i is affected by many factors, so that it cannot be represented by a simple function. In order to solve this problem, we use regression to analyze the impact of vehicle density and vehicle velocity on the packet transmission time through an individual road section.

pv_{r_i} instead of pd_{r_i} is used as the response variable of the regression model, because pd_{r_i} might be infinite in a sparse vehicular network, yet using too extensive a value range for the response variable may seriously affect the effectiveness of the regression model. On the contrary, the value range of pv_{r_i} is limited: obviously $pv_{r_i} \geq 0$, and the upper bound appears when the vehicle density is so large that the data packet can be transmitted consistently without staying on the individual vehicles. Suppose the transmission range is R, the transmission delay of one hop is τ, then $0 \leq pv_{r_i} \leq R/\tau$.

According to the strict lower and upper bound of pv_{r_i}, a logistic regression model is used on pv_{r_i}:

$$E\{pv_{r_i}\} = \frac{R}{\tau} \cdot \frac{e^{(X^T \cdot \beta)}}{1 + e^{(X^T \cdot \beta)}}$$

where $X = [1, \rho^0_{r_i}, v^0_{r_i}, \rho^1_{r_i}, v^1_{r_i}]^T$ and $\beta = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_4]^T$.

In this regression model, the historical data is used to calculate the edge weights and to predict the future traffic patterns. The dataset SUVnet (see Section IV) is used to fit the value of the variable vector $\beta = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_4]^T$.

The result path has the shortest transmission delay.

Since TARA does not require future information and the individual vehicles can calculate the shortest path on their own, TARA is distributed and practical compared to MDHA.

C. Routing rules

1) Path planning and updating: After the computation of the edge weights, TARA uses the *Dijkstra* algorithm to find the minimum cost path between the routing source and the destination. The packet is transmitted from one road intersection towards the next along this path.

The edge weights in VTW and RW vary over time. TARA sets a time threshold ΔT, updates the edge weights in ΔT, and recalculates the minimum shortest path where the routing source is reset to the current position of the data packet.

2) Local transmission strategy: The local transmission is between two adjacent road intersections, where geographic routing is used. Since the destination in the local transmission is always the endpoint of the road, the local minimum point problem in geographic routing can be effectively reduced.

3) Recovery mechanism: Since the real transmission depends on the individual vehicles and the vehicles may not be in full accordance with the planned path, we define the allowed area around the planned path:

- The central axis of the allowed area is coincident with the central axis of the road section.
- The width of the allowed area is in multiples of the transmission range.

The recovery is then executed as follows:

- If the packet deviates from the planned path but is still inside the allowed area, no correction.
- If the packet is outside the allowed area, TARA calculates the shortest path between the data packet and the allowed area, and uses geographic routing to lead the packet back into the allowed area.

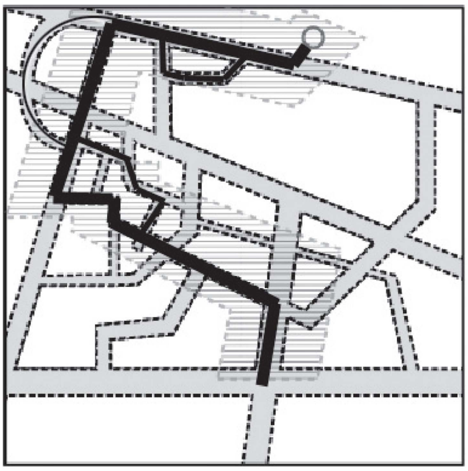


Fig. 2. Packets in the allowed area are acceptable

For example, the thickest line in Fig. 2 is the planned path. If the data packet is on the line with the medium thickness, TARA will make no correction since the line is inside the allowed area (indicated by the shadow area). However, if the data packet is on the thinnest line, TARA will first pull the packet back into the allowed area.

IV. SIMULATION

A. Simulation dataset

The simulation is based on the dataset of the Shanghai Urban Vehicle Net (SUVnet)[15], which collected real-time GPS data of over 4000 taxis from the year 2006 to 2007 in the urban area of Shanghai (about 102 square kilometers).

The original GPS records contain the taxi ID, the timestamp, the vehicle velocity, the driving direction (the geographical angle clockwise from the north) and the vehicle location (the longitude and latitude coordinates).

B. Data preprocessing

The original GPS data is coarse-grained (the time interval between two GPS records is 40 seconds) and imprecise (the GPS coordinates may deviate from the road). As a consequence, data preprocessing is needed to better shape the original GPS data for the simulation.

1) Map matching: In order to map the GPS data onto a digital road map of Shanghai (using the Geographic Information System repository), we choose the candidate roads where:

- The distance between the road and the taxi is less than 40m.
- The directional difference between the road and the taxi is less than 45^o.

If one of the candidate roads is the same as the last matched road, the last matched road is chosen; otherwise, the taxi is mapped onto the nearest candidate road.

Fig. 3. The selection of routing destinations

2) Route determination: The taxi route between two consecutive GPS data points is calculated following the principle of minimizing turns: The taxi will turn onto another road only if it cannot get nearer to the destination at the next road intersection along the current road.

3) Data interpolation: Given the length s (meters) of the route and the time interval t (seconds) between two consecutive GPS data points, intermediate data are generated for every s/t meters, so that the time interval between two interpolated data points is one second.

C. Simulation configuration

The parameters in the simulation are configured as follows:

1) Routing start time: Ten o'clock in the morning of five consecutive days beginning from 2006.11.30.

2) Routing source: For each routing start time, randomly choose one hundred vehicles in the central area of the city.

3) Routing destination: 532 grid intersections with fixed positions, as shown in Fig. 3. The horizontal and vertical distances between the adjacent intersections are both 500m.

4) Longest routing time: If the destination has not been reached in two hours, the routing is marked as failed.

5) Transmission range: 100m.

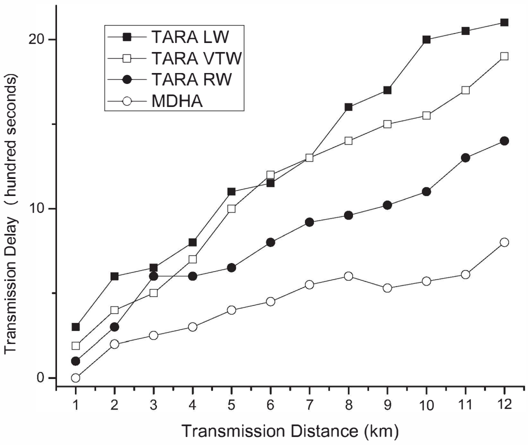
6) The width of the allowed area: 500m.

Based on the simulation configurations, we build a simulator to implement TARA.

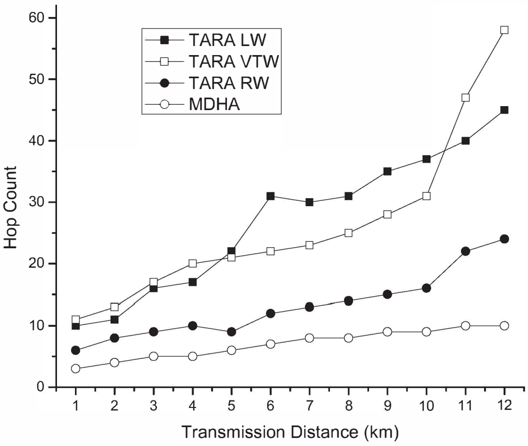
For each source and destination pair, the minimum cost path is first calculated using LW, VTW and RW. For every five minutes, the minimum cost path is recalculated with the latest GPS data, where the routing source is reset to the current location of the data packet.

For each timestamp, the vehicle currently carrying the packet will check its location and discover its neighbors within the transmission range. It will forward the packet to its neighbor that is nearest to the destination on the minimum cost path among all the neighbors.

If the vehicle cannot find such a neighbor and is still in the allowed area, it will keep carrying the packet. If the vehicle is outside the allowed area, it will forward the packet to its neighbor that is geographically nearest to the planned path.



(a) Delay



(b) Hop count

Fig. 4. Comparison of the three edge weighting methods

D. Simulation results

The delivery rate, the transmission delay, and the hop count to reach the destination are used as the simulation metrics.

Firstly we simulate TARA with the three different edge weighting methods separately and compare their performance.

As can be seen from Fig. 4 and Fig. 5, the regression weighting method achieves much better performance than the other two in all metrics. Thus we use TARA with RW for further algorithm comparisons with GPSR and geographic routing. MDHA is also included to indicate the optimal upper bound of the routing performance.

As can be seen from Fig. 6, GEO routing shows a performance closer to the optimal solution for short-distance data transmissions. However, as the transmission distance increases, the performance gap grows dramatically larger. One possible reason is that the GEO routing family is greatly influenced by

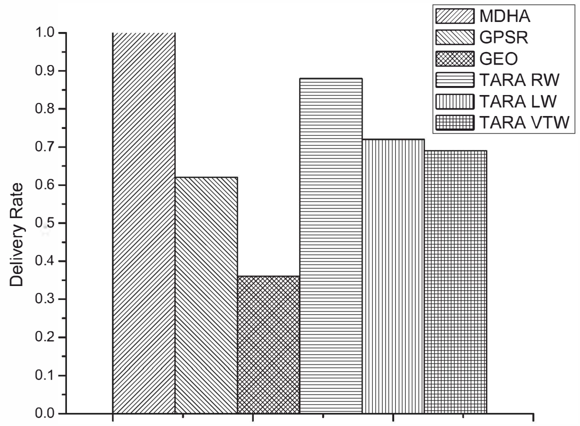


Fig. 5. Delivery rates

the local minimum point problem. On the contrary, TARA limits GEO routing to the routing on the road sections, where the routing target is always the road endpoint. As a consequence, the local minimum point problem is greatly reduced, which leads to a better performance.

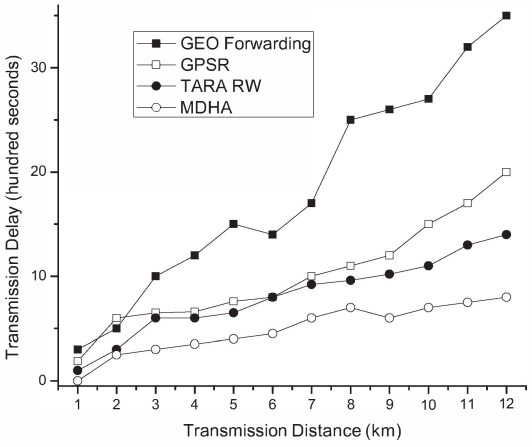
V. FUTURE WORK

It is possible that a vehicle cannot find a proper neighbor to forward the packet (especially at road intersections), which may lead to routing failures. In general, there are two ways to solve this problem: 1) Recalculate the minimum cost path according to the new position of the vehicle; 2) Keep replicas along the road section, so that the packet can still be forwarded along the pre-defined path even if a subset of the vehicles carrying the data copies fail to find the appropriate neighbors. However, path recalculation may affect the routing efficiency, while replicas will introduce additional communication and management overheads. We would like to further analyze and implement these two approaches in the future and compare their performance with the original TARA.

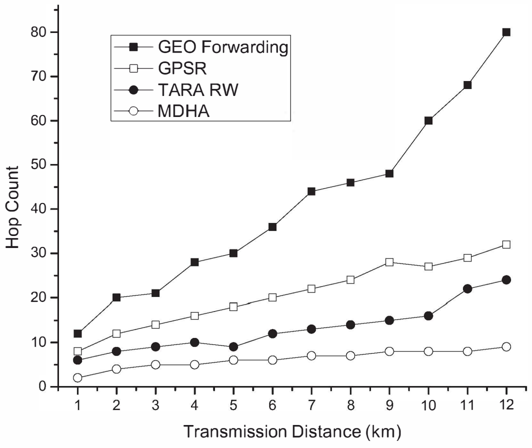
Currently the simulation chooses 10:00am as the routing start time. In order to further evaluate TARA, more diversified routing start times (include peak and non-peak time) will be chosen for the follow-up simulations.

REFERENCES

[1] K. Fall, "A delay-tolerant network architecture for challenged internets," *ACM SIGCOMM*, 2003.
[2] K. Fall and S. Farrell, "DTN: an Architectural Retrospective," *IEEE Journal on Selected Areas in Communications*, vol. 26, 2008.
[3] S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," *ACM SIGCOMM*, 2004.
[4] A. Vahdat and D. Becker, "Epidemic routing for partially-connected ad hoc networks," *Tech. Report*, 2000.
[5] E. Bulut, "Opportunistic Routing Algorithms in Delay Tolerant Networks," 2011.
[6] J. Leguay, T. Friedman, and V. Conan, "DTN routing in a mobility pattern space," *ACM SIGCOMM*, 2003.



(a) Delay



(b) Hop count

Fig. 6. Comparison of TARA RW, GPSR and GEO forwarding

[7] A. Chaintreau, J. Scott, and R. et al, "Pocket switched networks and human mobility in conference environments," *ACM WDTN*, 2005.
[8] C. Lorchert, M. Mauve, and Holger, "Geographic Routing in City Scenarios," *ACM SIGMOBILE*, 2005.
[9] B. Carp and H. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks," *ACM MOBICOM*, 2000.
[10] K. Lee, "VCLCR: a Piratical Geographic Routing Protocol in Urban Scenarios," *Tech. Report*, 2008.
[11] F. Kuhn, "Geometric Ad Hoc Routing: of Theory and Practice," *ACM PODC*, 2003.
[12] Z. Chen, "Ad hoc relay wireless networks over moving vehicles on highways," *ACM MOBIHOC*, 2001.
[13] S. Pallottino and M. Scutella, "Shortest Path Algorithms in Transportation Models: Classical and Innovative Aspects," *Equilibrium and Advanced Transportation Modelling*, 1997.
[14] R. Marler and J. Arora, "Survey of Multi-Objective Optimization Methods for Engineering," *Structural and Multidisciplinary Optimization*, vol. 26, 2004.
[15] H. Huang, P. Luo, and M. L. et al, "Performance Evaluation of SUVnet with Real-Time Traffic Data," *IEEE Trans. Vehicular Technology*, vol. 56, pp. 3381–3396, November 2007.