

START: Status and Region Aware Taxi Mobility Model for Urban Vehicular Networks

Haiquan Wang
 Wenjing Yang
 Jingtiao Zhang
 Jiejie Zhao

School of Software, Beihang University, Beijing, P.R.China
Beijing Key Laboratory of Network Technology, Beijing, P.R.China
 Yu Wang*

Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC, USA

Abstract

Using a realistic mobility model will enhance the validity of simulations. However, the difficulty lies in discovering laws from large amounts of data and applying those rules. Researchers have been working on mobility model extracting from real data set, whereas the taxi behavior differences between different taxi statuses have been ignored in previous works. Based on the experience in daily life, two assumptions related to the taxi status, one is the behavior of taxi will be influenced by the statuses and the other one is the macroscopic movement is related with different geographic feathers in corresponding status, are introduced and estimated by actual data. Based on the two assumptions, a novel taxi movement model START is proposed with respect to taxi statuses. The simulation results illustrate that the proposed movement model has a good approximation with reality in trace samples, distribution of nodes and the contact characteristics.

keyword

mobility model, taxi status, region recognition

I. Introduction

In vehicle ad hoc networks (VANETs) [1], realistic movement model is an important way to improve route planning, control traffic situations, or solve the vehicle-to-vehicle communication problems. However, movement models will influence simulation performance,

*This research has been partially supported by the US National Science Foundation (NSF) under Grant No. CNS-1319915 and CNS-1343355, the National Natural Science Foundation of China (NSFC) under Grant No.61300173 and No. 61170295, the Project of Aeronautical Science Foundation of China under Grant No.2013ZC51026 and No.2011ZC51024, the Fundamental Research Funds for the Central Universities under Grant No. YWF-12-LXGY-001, and the State Key Laboratory Software Development Environment and Network Information and Computing Center of Beihang University.

since movement model defines the nodal movement pattern including speed and direction. Whereas, large amounts of data are difficult to utilized directly. It is necessary to work on realistic mobility models. Some researchers [2], [3] modeled the vehicular movement, extracting different feathers from real data sets. But taxi status is ignored in the previous works. In this work, a STatus and Region Aware Taxi mobility model, START, is proposed based on the real taxi GPS data. Two assumptions are introduced in section II. We assume that the taxi behavior and geographic feathers are related with different status. They are validated to be reasonable by the statistical analysis of the data set. START is modeled based on these assumptions. In the macro scope, instead of simply dividing the area into coarse-grain regions, we divide the area into two set of regions according to the passenger load or drop event density. When a taxi take a passenger, the current region will be selected in the region set of load-event and the destination region, where the drop-event happens, will be selected in the region set of drop-event. We investigate the relationship between load-event regions and drop-event regions. Pathes from the sources to destinations will be found by Dijkstra algorithm. For microscope, the *speed* for the two taxi statuses are discussed respectively. Simulations are carried out to compare the similarity of node trace characteristics and contact characteristics. The results show that our mobility model has a good approximation with the real scenario. The rest of our paper is organized as follows: Section II proposes two assumptions which are further validated by statistical results of real data. Section III presents the modeling process. Simulation results are demonstrated in Section IV. Finally, Section V concludes this paper.

TABLE I
EXPLANATION OF EVENTS AND STATUS

Event	Explanation
0(drop)	A taxi's status change to vacant.
1(load)	A taxi's status change to occupied.
2	Set up defense.
3	Cancel defense.
4	No event happened.
Status	Explanation
0(vacant)	A taxi is vacant.
1(occupied)	A taxi is occupied.
2	A taxi is setting up defense.
3	Stop running.

II. Assumptions and Statical analysis of Taxi Trace

In this section, we mainly focus on statistical analysis on the speed and duration characteristics on the data set. Firstly, the data set will be introduced in section II-A. Then, two assumptions are proposed and validated in the following sections.

A. Trace Dataset: Beijing Taxi Traces

A real-world GPS data set is used in this paper, which was generated by 12,455 taxis in Beijing, China within 5 days from March 3rd,2011 to March 7th,2011. This data set has been used by several key researches and application programs of Intelligent Transportation Systems (ITS) in Beijing, China. The number of participated taxis (12,455 taxi node) is 18% of the total taxis in the city. Each taxi is equipped with a GPS device and upload its information (including location, speed, direction) about every 60 seconds. There are about 1.22×10^8 records in total. Each row includes a base station ID, company name, taxi ID (*id*), timestamp (*t*), current location (*l*, including longitude and latitude), location of 54 format, speed, acceleration, status of the taxi, event, and height. Of all the fields, taxi ID, timestamp, and current location, status and Event are used in this paper. Note that GPS traces from taxis have been used recently for inferring human mobility [14] and modeling city-scale traffics [15]. Therefore, we believe that they are suitable to characterize the contact patterns among vehicles in large-scale urban scenario.

Especially, there are five types of events and four types of status. The explanations are as table I. We only discuss the vacant status and occupied status in this paper. Accordingly we utilize the load-event and drop-event as well.

B. Assumptions

According to the experience in daily life, the following two assumptions are given:

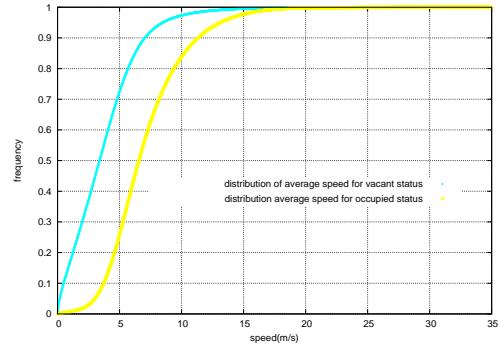


Fig. 1. Speed Distribution for vacant and occupied status.

- **Assumption 1:** The behavior of a taxi will change when its status changes. When a taxi is taking passengers, its destination is fixed, and the speed of it is relatively faster. In contrast, when a taxi is vacant, it will slow down or even stop to search for potential passengers along the road. Thus, taxi behavior characteristics, such as speed and the status duration varies consequently.
- **Assumption 2:** The movement behavior of taxies associates with geographic feathers.
 - 1) The destination selection will be influenced by different regions.
 - 2) events occurs in different regions un-evenly, passenger drop and on events are distinct.

Therefore, we analysis the speed,duration and passenger load/drop events distribution to estimate our assumptions.

C. Speed analysis

In this section, we investigate the average speed *speed* in each status. For example, taxi *i* drives in occupied status for a distance *d* using time *t*, then the average speed in this status is *d/t*. From March 3th to 7th, 2011, the *speed_{empty}* = 3.627m/s, while that for occupied status is *speed_{occupied}* = 7.083m/s.

To further investigate the speed distribution, proportion for every *speed* section is calculated, in fig.1. For example, dot(x,y) means 2.45% records fall in the speed range [0,20]km/h. Fig. 1 shows that *speed* distribution differs for each status. For vacant status, the *speed* gather together at 1 also demonstrates that the speed distribution is with strong regularity for each status.

D. Status duration analysis

The duration distribution for each status are shown in fig.. Status duration represents the time length of a taxi staying in a certain status. The red line presents the duration time distribution for vacant status, and the

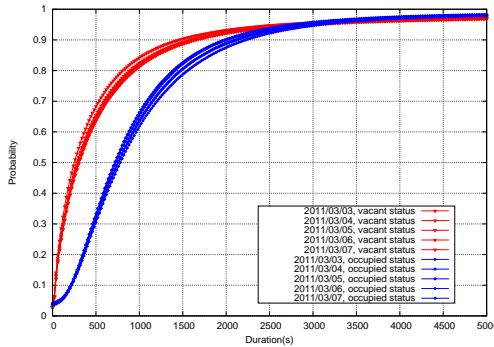


Fig. 2. Status duration distribution.

green one is for status 1. A dot of the line means the proportion of the duration. A peak exists in each line, and it is obvious that the peak of the red line is earlier than that of the other line. And the value of duration for status 0 trends to be smaller. It accords to the realistic situation, because drivers tend to shorten the waiting time to raise their income.

The statistical results are consistent with the *assumption 1*, that is, the behaviors of taxis are similar within each status while it differs between the two statuses.

E. Taxi event distribution

To validate *assumption2*, we quantitatively analyze vehicles density in one hour. By dividing the whole network into $100 * 100$ grids, taxi density distributions for event 0, 1 are computed in each cell.

Figure.3 shows the on/off events distribution in three time sections. In the morning, people begin to get out. so the on-event distribute is more even than that of off-event. this phenomenon may be caused by the on-event spots are mainly at the homes of the citizens, while the off-event spots tend to gather together at workplaces, railway stations or scenic spots. Besides, in the evening, the dispersion of on-event is of lower degree than that of off-event, for people coming home at that time.

The amount of loading passengers in each cell shows geographic feathers:the distribution is uneven, and the difference between on/off-event distributions illustrates the on/off-event regions are different which support the *assumption 2*.

III. Modeling

Movement model defines the mobility pattern of nodes, which can be represented as a collection of paths $\text{Paths} := \langle p_1, p_2, \dots, p_n \rangle$, so dose START. A p_i includes two steps, destination selection and moving process from source location to destination.

Destination selection: In START, to select a destination of a node is closely related to note only its current location but also its current status. Dividing the area into regions by the density of passenger load/drop events or loading passenger events respectively, two transition probability matrixes are calculated, one is the probability from passenger drop event regions $\{\text{REGION}_{m,\text{drop}}\}$ to passenger on events regions $\{\text{REGION}_{n,\text{load}}\}$. Note that $\bigcup\{\text{REGION}_{m,\text{drop}}\} = \bigcup\{\text{REGION}_{n,\text{load}}\} = \text{AREA}$. If the status of a taxi changes to vacant, its current location determine a $\text{REGION}_{i,\text{drop}}$. Consequently, a destination region in $\{\text{REGION}_{n,\text{load}}\}$ will be selected by querying the transition matrix from $\{\text{Region}_{m,\text{drop}}\}$ to $\{\text{Region}_{n,\text{on}}\}$. Then, START will randomly select a map node in the region as the destination. As to the status of a taxi changing to Occupied, the destination selection process is similar. During this process, the region transition matrix will be utilized according to the current status.

Moving process: When the source location (current location) and destination location is given, next step is to find a path and set the speed. To simplify, we adopt the Dijkstra algorithm ,which will find a shortest path from source to the destination, to route on map. The speed is assigned by the average speed distribution of the corresponding status.

Based on the design above, we model the movement on the speed, duration and region transition matrix respectively.

A. Region transition probability

A travel path of a taxi can be simplified as a multi-hop process, in which a hop indicates an load/drop event happened. Seeing that, we define a *region transition probability* to figure out the probability of the next hop falling in a definite region j from the current region i . Particularly, two successive events are different. Likewise, the region i and j are recognized by different metrics, that is, drop or load event distribution. It is more reasonable. For an instance, if the taxi is occupied, the next hop event is the drop one. Hence, choosing a target region from a region set divided by drop event distribution is more logical.

To calculate the region transition probability, the **region recognition process** should be executed in advance.

Firstly, we divide the area into 100×100 grids, and define cells in it as equation 1. Then, we consider region as adjacent cells as equation 2.

$$\text{CELL}_{x,y} ::= \{(lon, lat) | x \leq \frac{lon}{len_x} < x + 1, y \leq \frac{lat}{len_y} < y + 1\}, \quad (1)$$

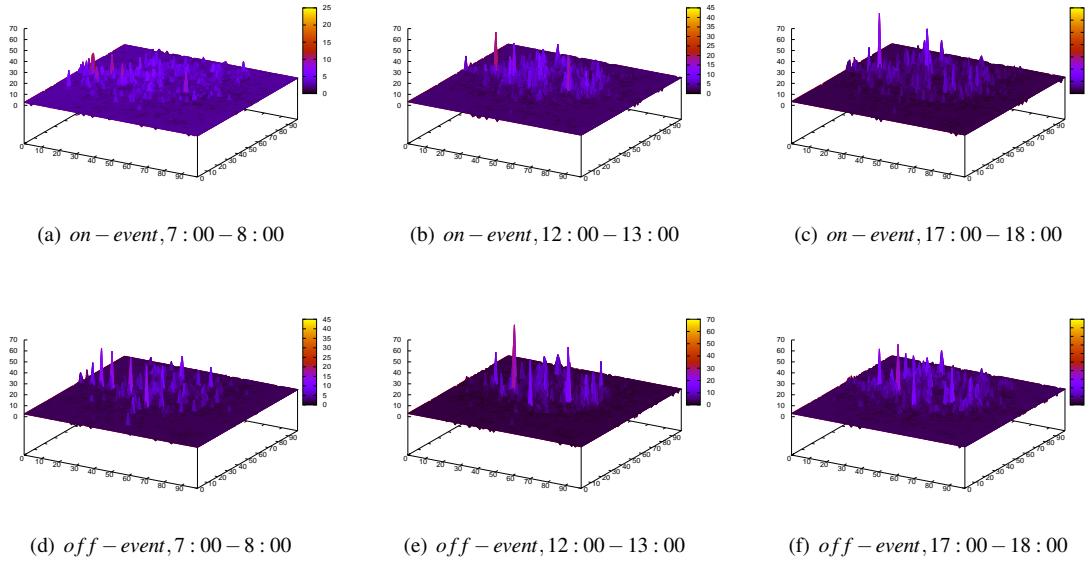


Fig. 3. Taxi density for on-event and off-event in one hour

$$\begin{aligned} REGION_m ::= & \{CELL_{x,y} | \exists CELL_{i,j} \in REGION_m \\ & \Rightarrow \|x - i\| \leq 1, \|y - j\| \leq 1\} \end{aligned} \quad (2)$$

By clustering cells into regions, two set of region $\{REGION_m^{load}\}$ and $\{REGION_n^{drop}\}$ can be recognized. The clustering idea is to put adjacent cells with event density larger than the event threshold η into one cluster. To avoid the size of the region become too large or too small, we set a $CLUSTERSIZE$ to restrict a region, say $\|REGION_i\| \leq CLUSTERSIZE$, and only limit the top 200 regions, in which $CELL_{x,y}.events \geq \eta$. After that, the other cells not belong to the top 200 regions, will be cluster into regions, while $\|REGION_j\| \leq CLUSTERSIZE$. We sort the cells by event density in descending order, and begin with the first cell to search its neighbors whether to join the same region using breadth traversal. The region recognition results for load/drop events are shown in figure 4. The detail clustering algorithm is presented in the appendix. In addition, the $CLUSTERSIZE = 200$, $\eta = 121$ for on event and $\eta = 141$ for drop event set by the average event density of the top 5000 cells order by its event density.

The calculate process of the region transition probability: After clustering cells into regions, the transition probability from $REGION_i^{load/drop}$ to $REGION_j^{drop/load}$, denoted as $p_{i \rightarrow j}^{load \rightarrow drop/drop \rightarrow load}$. To substantiate, the calculate process of $p_{i \rightarrow j}^{load \rightarrow drop}$ will be introduced in detail. Clearly, the records in $REGION_i^{load}$ can be required from the data

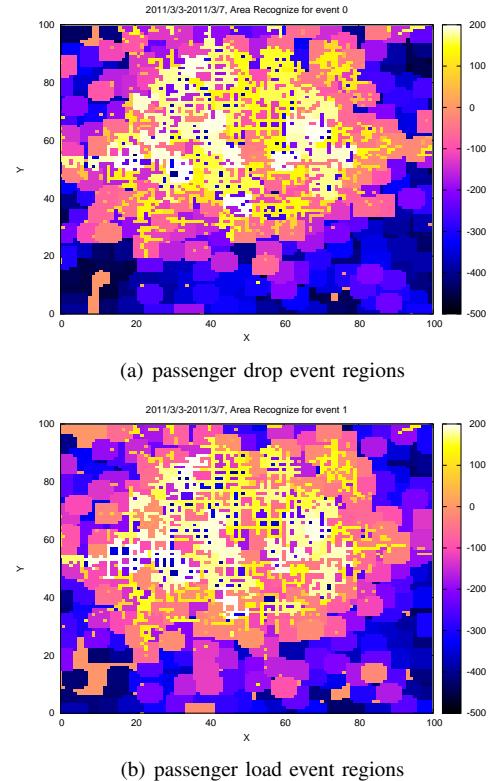


Fig. 4. Region recognition

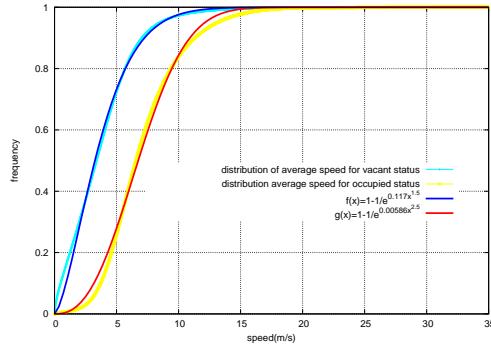


Fig. 5. The fit result of the cumulative speed distributions.

set easily. the record amount is denoted as $\|RECORDS_i^{load}\| = \{record | record.location \in REGION_i^{load} \cap record.event = load\}\|$. For $record \in RECORD_i^{load}$, the next event and location can be easily required. Therefore, the record can be associated with the its next hop information to $(taxiid, location_{current}, event, event_{next}, location_{next})$. The $RECORDS_{i \rightarrow j}^{load \rightarrow drop} = \{record | event = load \cap event_{next} = drop \cap location_{current} \in REGION_i^{load} \cap location_{next} \in REGION_j^{drop}\}$ will be obtain.

$$p_{i \rightarrow j}^{load \rightarrow drop} = \frac{\|RECORDS_{i \rightarrow j}^{load \rightarrow drop}\|}{\|RECORDS_i^{load}\|} \quad (3)$$

$$P^{load \rightarrow drop} = (p_{i \rightarrow j}^{load \rightarrow drop})_{m \times n} \quad (4)$$

$$P^{drop \rightarrow load} = (p_{i \rightarrow j}^{drop \rightarrow load})_{n \times m} \quad (5)$$

B. Parameter estimation of speed distribution

In this section, we modeling the speed distribution. we fit the cumulative status average speed distribution to get the cumulative probability distribution function, and then take a derivative with it to obtain the speed probability distribution.

From figure. 5, the average speed distribution shows exponential law. Given that, we set the function form as follows:

$$\begin{cases} f(x) = 1 - \exp(-a_1 x^{1.5}) \\ g(x) = 1 - \exp(-a_2 x^{2.5}) \end{cases} \quad (6)$$

The fit formulas are given as formulas 6. $f(x)$ is the function form for the average speed distribution of vacant status, and the other one $g(x)$ is for that of occupied status.

Fig. 5 also plots the fit results.

The rms of residuals for each fit are as table II. The smaller rms of residuals means better fitting. In the

TABLE II
THE PARAMETER AND THE RMS OF RESIDUALS OF FITTING CURVES

Categories	rms of residuals	
$f(x)$	$a1 = 0.117$	0.0113159
$g(x)$	$a2 = 0.0586$	0.0137029

table, the values are all less than 0.02, showing good similarity.

IV. Model Verification

In this section, START mobility model is validated on the aspects of node distribution and contact characteristics. All movement models are implemented on Opportunistic Networking Environment (ONE)[16].

Shortest Path (SP) movement model based on the map in Beijing is an other comparison, which is implemented by ONE. It also moves along the map roads by Dijkstra algorithm. The RWP model is another comparison, because it is proved to be an efficient model modeling the nodal movement in VANETs. But it takes no consideration of the node statuses and geographical distribution.

The START, SP and RWP mobility model are compared with the real trace. In simulations, Node number is set as 4000 and scenario in area $24445 * 23584m^2$ (a sub-map of the whole area), including fourth ring roads in Beijing. The simulation time is three hours and the warm up time for reports is one hour, so that the nodal movement and position will not be affected by its initial position. The communication range is 200m.

A. Traces and distribution of nodes

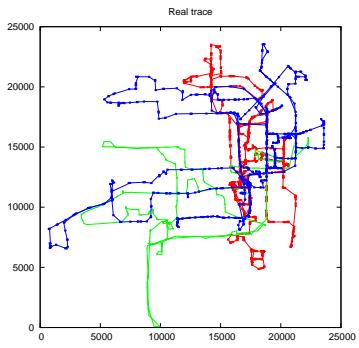
Trace samples and their snapshots are demonstrated in this section, shown as fig. 6 and fig.7.

Fig.6 shows the trace in one day. The trace of the real data and START only cover some part of the area, while the trace of SP and RWP almost go through the whole area. Because SP and RWP will select a destination randomly, while START takes the associations between current region and destinations into consideration which satisfies the movement rules of taxies.

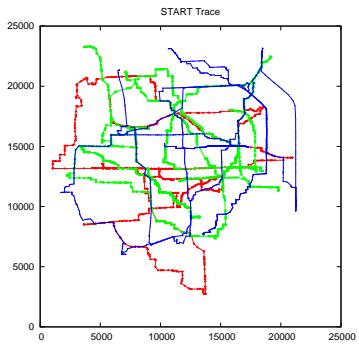
From fig.7, Real trace START and SP movement model exhibit the road structure. However, the node distribution of RWP is much uniform.

B. Contacts characteristics

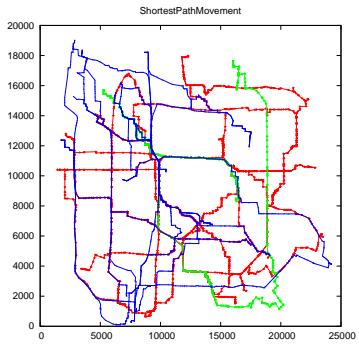
The contact time and inter contact time are evaluated as the indicators to validate the similarity. The speed range of RWP and SP need to be configured. To ensure the accuracy, we choose three speed range



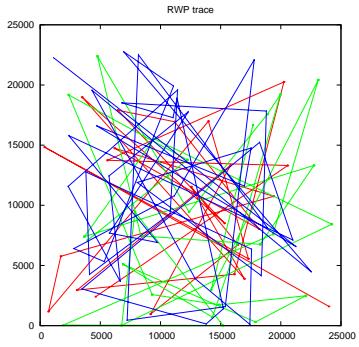
(a) Real Trace



(b) START

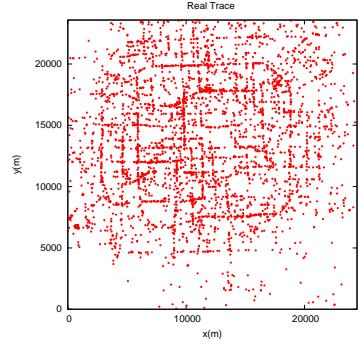


(c) SP

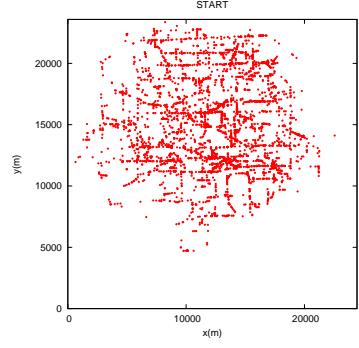


(d) RWP

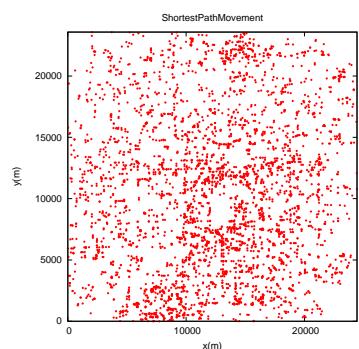
Fig. 6. Trace samples



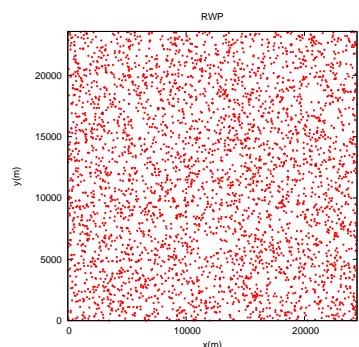
(a) Real Trace



(b) START



(c) SP



(d) RWP

Fig. 7. Trace snapshots

[0, 44.4]m/s, [0, 33.3]m/s and [0, 22.2]m/s, that is, the upper bounds of speed are 80, 120, 160 km/h.⁷

Fig. 8. (a)(b) is the contact time and inter-contact distribution, which shows the probability of the contact or inter-contact time smaller than certain time length. To substantiate, a point (200, 0.5) in these figures means the probability is 0.5 when contact or inter-contact time is shorter than 200s. In addition, a point (x, y) in 8.(c) means the sum of contact time is y when the simulation time is x . In this figure, the three green lines of SP and the three blue lines of RWP are coincide with each other. To recognize the differences, we set y axis as log-scale in fig.8.(d). The curves of the total contact time vs. time presents a liner law. For SP and RWP, the differences of speed ranges show little influence on these curves. Clearly, the rank of the contact characteristic similarity with the real data is START > SP > RWP.

To conclude, by comparing the node distribution and contact characteristics, the START mobility model performs good similarity with the real data. The evaluation results conform to our expectations. Because START takes use of speed and geographic feathers related with status, While SP employs the map information and RWP is a random model taking use of no realistic data.

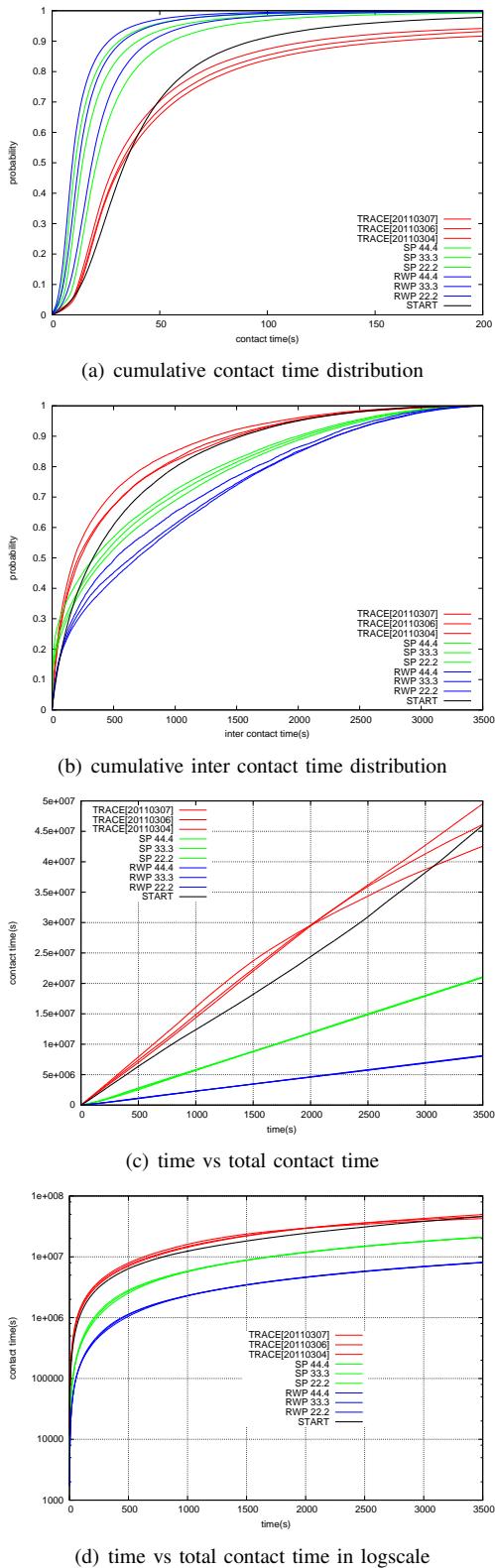


Fig. 8. Contact times distribution and Cumulative ICT distribution

V. Conclusion

Since the mobility model is essential for mobile network, a novel mobility model START based on real GPS trace data is proposed. By assuming the taxi behavior is related with its statuses and geographic feathers, statistical experiments are conducted to verify those assumptions using the real trace data. Further, its parameter—average speed of each status are estimated respectively, and the region transition probability is calculated. In this case, macroscopic movement, a node moves switch between load-event regions and drop-event regions, and microscopic movement(speed for each status) can be defined. Finally, the START is implemented on ONE simulator and estimate it by comparing with the real trace, RWP and Shortest Path movement Model. Comparing the node distribution and contact feathers, START shows better performance. Simulation results demonstrate that START has a good approximation with reality.

References

- [1] S. Yousefi, M. Mousavi, and M. Fathy, "Vehicular ad hoc networks (vanets): Challenges and perspectives," in *ITS Telecommunications Proceedings, 2006 6th International Conference on*, June 2006, pp. 761–766.

- [2] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," in *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, vol. 6, 2006, pp. 1–13.
- [3] H. Huang, Y. Zhu, X. Li, M. Li, and M.-Y. Wu, "Meta: A mobility model of metropolitan taxis extracted from gps traces," in *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*, 2010, pp. 1–6.
- [4] X. Lu, Y.-c. Chen, I. Leung, Z. Xiong, and P. Li O, "A novel mobility model from a heterogeneous military manet trace," in *Ad-hoc, Mobile and Wireless Networks*, ser. Ad-hoc, Mobile and Wireless Networks. Springer, 2008, pp. 463–474.
- [5] S. Ahmed, G. C. Karmakar, and J. Kamruzzaman, "An environment-aware mobility model for wireless ad hoc network," *Computer Networks*, vol. 54, no. 9, pp. 1470–1489, 2010.
- [6] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," in *Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking*. ACM, 1998, pp. 85–97.
- [7] A. K. Saha and D. B. Johnson, "Modeling mobility for vehicular ad-hoc networks," in *Proceedings of the 1st ACM International Workshop on Vehicular Ad Hoc Networks*, 2004, pp. 91–92.
- [8] C. P. Mayer and O. P. Waldhorst, "On the impact of graph structure on mobility in opportunistic mobile networks," in *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, 2011, pp. 882–887.
- [9] W. Peng, G. Dong, K. Yang, J. Su, and J. Wu, "A random road network model for mobility modeling in mobile delay-tolerant networks," in *Mobile Ad-hoc and Sensor Networks (MSN), 2012 Eighth International Conference on*, 2012, pp. 140–146.
- [10] F. J. Martinez, J.-C. Cano, C. T. Calafate, and P. Manzoni, "Citymob: a mobility model pattern generator for vanets," in *Communications Workshops, 2008. ICC Workshops '08. IEEE International Conference on*, 2008, pp. 370–374.
- [11] D. R. Choffnes and F. A. N. E. Bustamante, "An integrated mobility and traffic model for vehicular wireless networks," in *Proceedings of the 2nd ACM international workshop on Vehicular ad hoc networks*, 2005, pp. 69–78.
- [12] A. Mahajan, N. Potnis, K. Gopalan, and A. Wang, "Modeling vanet deployment in urban settings," in *Proceedings of the 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, 2007, pp. 151–158.
- [13] H. Huang, D. Zhang, Y. Zhu, M. Li, and M.-Y. Wu, "A metropolitan taxi mobility model from real gps traces." *J. UCS*, vol. 18, no. 9, pp. 1072–1092, 2012.
- [14] R. Ganti, M. Srivatsa, A. Ranganathan, and J. Han, "Inferring human mobility patterns from taxicab location traces," in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '13. New York, NY, USA: ACM, 2013, pp. 459–468.
- [15] J. Aslam, S. Lim, X. Pan, and D. Rus, "City-scale traffic estimation from a roving sensor network," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, ser. SenSys '12. New York, NY, USA: ACM, 2012, pp. 141–154.
- [16] A. Keraen, J. Ott, and T. Karkkainen, "The ONE simulator for DTN protocol evaluation," in *Proceedings of the 2nd International Conference on Simulation Tools and Techniques*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009, p. 55.

Appendix

Algorithm 1 Clustering

Require: $Cells = \{CELL\}$

η ##events threshold

$CLUSTERSCALE$

$REGIONSEED = 200$

$ClusterQueue = \emptyset$

$UsedCells = \emptyset$

Sort $Cells$ by events DESC

for $CELL_{x,y} \in Cells$ **do**

if $CELL_{x,y} \notin UsedCells$ **then**

$CELL_{x,y}.region = REGIONSEED$

 size = 1

$CLUSTERSEED = CLUSTERSEED - 1$

$ClusterQueue.enqueue(CELL_{x,y})$

$UsedCells.add(CELL_{x,y})$

while $ClusterQueue \neq \emptyset$ **do**

$CELL_{x,y} = ClusterQueue.dequeue()$

if $REGIONSEED \geq 0$ **and** $CELL_{x,y}.events \geq \eta$ **then**

 enqueueNeighbor($CELL_{x-1,y}$)

 enqueueNeighbor($CELL_{x-1,y-1}$)

 enqueueNeighbor($CELL_{x-1,y+1}$)

 enqueueNeighbor($CELL_{x+1,y}$)

 enqueueNeighbor($CELL_{x+1,y-1}$)

 enqueueNeighbor($CELL_{x+1,y+1}$)

 enqueueNeighbor($CELL_{x,y-1}$)

 enqueueNeighbor($CELL_{x,y+1}$)

else

 enqueueNeighborOthers($CELL_{x-1,y}$)

 enqueueNeighborOthers($CELL_{x-1,y+1}$)

 enqueueNeighborOthers($CELL_{x-1,y-1}$)

 enqueueNeighborOthers($CELL_{x+1,y}$)

 enqueueNeighborOthers($CELL_{x+1,y-1}$)

 enqueueNeighborOthers($CELL_{x,y-1}$)

 enqueueNeighborOthers($CELL_{x,y+1}$)

Algorithm 2 enqueueNeighbor($CELL_{x,y}$)

if $CELL_{x,y}.events \geq \eta$ **and** size < $CLUSTERSCALE$ **and** $CELL_{x,y} \notin UsedCells$ **then**

$ClusterQueue.enqueue(CELL_{x,y})$

$CELL_{x,y}.region = REGIONSEED$

$UsedCells.add(CELL_{x,y})$

 size = size + 1

Algorithm 3 enqueueNeighborOthers($CELL_{x,y}$)

if size < $CLUSTERSCALE$ **and** $CELL_{x,y} \notin UsedCells$ **then**

$ClusterQueue.enqueue(CELL_{x,y})$

$CELL_{x,y}.region = REGIONSEED$

$UsedCells.add(CELL_{x,y})$

 size = size + 1
