

RESEARCH

Open Access

Empirical study on spatial and temporal features for vehicular wireless communications

Yingwen Chen^{1*}, Ming Xu¹, Yu Gu², Pei Li³, Lei Shi¹ and Xiaoqiang Xiao¹

Abstract

Static topology analysis is not sufficient for the dynamic vehicular *ad hoc* network. Understanding the evolving topology of vehicular *ad hoc* networkings (VANETs) caused by vehicle mobility is very important for routing protocol design and algorithm optimization. This paper explores the spatial and temporal features of vehicular network topologies based on two real taxi-trace datasets. The analysis results reveal that the whole topology of VANETs consists of a large number of small-sized connected components. Two quantitative metrics are proposed to measure the stability and location dependency of the connected components. When the communication range is greater than a threshold, a large proportion of vehicles will connect to the biggest connected component, which is relatively stable and covers the most part of the downtown region of the city. Based on the analytical results, we propose several design philosophies and new research issues for VANETs.

Keywords: Wireless communications; Spatial temporal analysis; Connected component; VANETs

1 Introduction

Vehicular *ad hoc* networking (VANET) is one kind of new technology supposed to provide innovative services for intelligent transportation systems. Equipped with dedicated short-range communications radios, vehicles can not only exchange messages directly with vicinity nodes but also communicate with other nodes through a number of intermediate nodes if all of them are connected. Since the nodes have the capability of communicating with each other, VANETs can provide a number of potential applications with highly diverse requirements. The three major classes of applications possible in VANETs are safety oriented, convenience oriented, and commercial oriented [1]. Safety applications include immediate collision warning, forward obstacle detection and avoidance, emergency message dissemination, and so on. Convenience applications can provide route maps with real-time traffic jams and accident conditions to help drivers find the shortest path in terms of time consumption. Commercial applications can provide internet access as well as communications between passengers in cars in the same

vicinity, allowing the passengers to surf the internet, watch online movies, and even play games.

To support information diffusion in VANETs, two kinds of network architectures are proposed. One is called vehicle-to-vehicle (V2V), in which vehicles can communicate when they are located in the communication range of each other. The other one is called vehicle-to-roadside (V2R) or vehicle-to-infrastructure (V2I) or vehicle-to-wayside (V2W), which is a cellular-like system. The roadside infrastructure is used as a basestation, and it may dominate the communication in its communication range. In V2V VANETs, vehicles are equal and information diffusion can be achieved by adopting MANET routing protocols. However, it is well known that moving vehicles result in a disconnected network topology [2]; thus, different kinds of carry-and-forward strategies [2-4] have been proposed to support intermittent connected networks. The carry-and-forward strategy is effective but not efficient, because the source might delay the forwarding until it meets the destination even though there is a routing path between them. It is shown that the packet forwarding delay caused by carry-and-forward can be several orders-of-magnitude longer than that caused by multi-hop forwarding over a connected network [5]. Consequently, we should figure out when and where the vehicles are connected and can be reached for

*Correspondence: cswchen@gmail.com

¹ College of Computer, National University of Defense Technology, 410073 Changsha, Hunan, China

Full list of author information is available at the end of the article

routing purpose. In V2R VANETs, the roadside infrastructure is usually more powerful than vehicles. To reduce system construction cost and improve communication efficiency, the deployment of roadside infrastructures should also be well studied. Intuitively speaking, the roadside infrastructures are better set up at those spots with poor V2V connectivity. However, moving vehicles bring a lot of uncertainty for infrastructure deployment. No one knows whether the well-connected vehicles will in turn be disconnected in the next minute, whether the connectivities of vehicles are depending on their location, and whether there is any difference between on-peak hours and off-peak hours. Therefore, connectivity distribution and time variance of network topology should be examined.

The purpose of this paper is to reveal the deep understanding of spatial and temporal dynamics of VANETs based on two real taxi-trace datasets collected from San Francisco, USA and Shenzhen, China. The analysis results are supposed to provide new guidelines for VANET design and protocol optimization. The main contributions of this paper are as follows:

1. We find that by adopting a reasonable communication range, a large number of vehicles are connected as a main component of the whole network.
2. We propose a metric to measure the stability of the connected component for a dynamic topology. Based on this stability factor, we find that the biggest connected component of the VANET is relatively stable even though each individual vehicle keeps moving.
3. We analyze the location dependency of the biggest connected component. We find that the most part of the downtown region of the city can be covered by the connected component of the VANET.
4. According to the spatial and temporal features that we have found, we propose several design philosophies and new research issues for VANETs.

The rest of the paper is organized as follows: Section 2 briefly summarizes the related work, Section 3 describes the definitions used in spatial and temporal analysis for VANETs, and Section 4 introduces the two taxi-trace datasets and presents the spatial and temporal discoveries as well as their implications for VANET design and protocol optimization. Finally, Section 5 concludes the article.

2 Related work

With the broadcast feature of wireless channels, a VANET is always modeled by a unit disk graph, in which two vertices are connected if their distance is below a fixed

threshold [6,7]. By adopting this assumption, percolation theory can be used to analyze the connectivity of VANETs. For example, quantitative relationships among network connectivity, vehicle density, and transmission range are derived in [8]. When considering more realistic constraints of VANETs such as non-uniform and non-Poisson distributions, or non-ideal environments with fading/shadowing impact, Miorandi et al. [9] proposed an equivalent $GI|D|\infty$ queuing model to analyze the connectivity of one-dimensional *ad hoc* networks. Their results claim that no connectivity can be obtained in condition of channel randomness. Based on the equivalent $GI|D|\infty$ model, the node isolation probability and the average size of connected components can be estimated in a one-dimensional case. For one-dimensional VANETs, Zhuang et al. [10] also derived the exact expression for the average size of the connected components and their size distribution. Furthermore, to analyze message propagation in two-dimensional traffic networks, the authors in [10] extended their model to calculate lattice-connectivity probability for all the blocks in a district. In real traffic, most vehicles are travelling in a co-directional way. Abuelela et al. [3] found that co-directional traffic is inherently partitioned into connected components and provided an analytical expression of the expected size of those connected components. By exploring the co-directional feature and the existence of connected components, the authors in [3] also designed an Opportunistic Packet Relaying protocol (OPERA) for packet delivery over disconnected VANETs.

In recent years, complex network- and network-science-related concepts [11] are widely adopted in many research domains. Monteiro et al. [12] decomposed the synthetic dynamical topology of VANETs into snapshots and calculated the macroparameters of the network such as the node degree distribution, the clustering coefficients, the average shortest path length, and so on for each topology snapshot. Based on the abstracted information from network science, a new efficient broadcasting protocol called UV-CAST has been proposed. To further explore the dynamics of VANETs in a completely new way, more analysis [13,14] based on real and realistic simulated traces have been carried out in terms of many other kinds of complex network metrics. Literature [13] makes a thorough investigation of both microscale and macroscale metrics including node degree, lobby index, link duration, network diameter, closeness centrality, betweenness centrality, number of communities, clustering coefficients, and so on. Similar to [13], literature [14] conducts node-level and network-level analysis including node degree, network assortativity, betweenness centrality, and so on. Besides these, literature [14] also analyzes connected components' dynamics when the number of vehicles and communication ranges vary.

Generally speaking, literatures [8-10] explore VANETs' topology based on theoretical connectivity models. These models are useful in discussing the critical conditions or the connectivity boundary. But these theoretical models still rely on strong assumptions, which are very difficult to achieve in real cases. For example, the model in [8] requires a uniform distribution of the vehicles. However, it has been proved not to be true in real mobility scenarios [15]. Literatures [11-14] provide new methods to explore VANETs' topology based on network science models. Literatures [13] and [14] also analyze the evolving topology by addressing the time-variant size of the connected components. However, their investigation on the size of the connected components cannot correctly reveal the evolving topology, since the size of the connected components can remain the same while the vertices in the connected components might be completely changed. Different from the current work on VANETs' topology analysis, we conduct our research work based on two real taxi traces and propose a new metric to measure the stability of connected components and evaluate the location dependency of the largest connected component of the VANET.

3 Definitions

By adopting the UDG model, which is the intersection graph of equal-sized circles [16], VANET topology can be abstracted as a non-directional graph tagged with time stamps. For better explanation, we first introduce some annotations for the related definitions.

3.1 Network model

The traditional static graph model in describing a network is $G = \langle V, E \rangle$, where V represents the nodes and E represents the edges between the nodes. However, a VANET is dynamic and the topology is evolving due to the movement of the vehicles. Therefore, the topology of VANET can be expressed by a time-stamped graph $G(T) = \langle V, E, T \rangle$, where V represents all the vehicles, E represents the links between two vehicles, of which the Euclidean distance is smaller than the wireless communication range R , and T is the time stamps. In real cases, due to obstacles such as high buildings, two vehicles might not establish a link even though their distance is smaller than the wireless communication range. In this situation, every vehicle can explore its real neighbor list by exchanging beacon messages with its neighbors. Consequently, VANET topology can be figured out based on the neighbor lists of all vehicles.

3.2 Routing path

Routing in a network is just to find a path in a given topology. We firstly define the routing path in a simplified static case. In a given time stamp t , the topology of

VANET is $G(t) = \langle V_t, E_t, t \rangle$. There is a non-empty sub-graph of $G(t)$, denoted as $P(t) = \langle V'_t, E'_t, t \rangle$. Assume $V'_t \subseteq V, E'_t \subseteq E, n = |V'_t|, V'_t = \{v_{a1}, v_{a2}, \dots, v_{an}\}$. $P(t)$ is called a *path* if and only if there exists $\sigma : V'_t \rightarrow V'_t; \sigma(v_{ai}) = v_i (i = 1, \dots, n)$, s.t. $E'_t = \bigcup_{j=1}^{n-1} (v_j, v_{(j+1)})$. Note that v_1 and v_n are called the two ends of the path at time stamp t . The path from v_1 to v_n can be denoted as Equation 1.

$$P_{v_1}^{v_n}(t) = \left\{ v_1 \overset{t}{\longleftrightarrow} v_n \right\} \quad (1)$$

The *length* of the path is $|V'_t|$, which is n . Note that there are probably more than one path from v_1 to v_n in $G(t)$. Therefore, the *distance* from v_1 to v_n is defined as the shortest path from v_1 to v_n in $G(t)$. Meanwhile, there might be more than one shortest path from v_1 to v_n in $G(t)$ as well.

In a general case, the routing path may pass through several time stamps. Without loss of generality, assuming the routing path from v_1 to v_n passes through a non-descending time-stamp set $T_s = \left\{ t_j \Big|_{j=1}^{j=m} \right\}$, we then have Equation 2.

$$P_{v_1}^{v_n}(T_s) = \bigcup_{j=1}^m \left(P_{v_{t_{j-1}}}^{v_{t_j}}(t_j) \right) = \bigcup_{j=1}^m \left\{ v_{t_{j-1}} \overset{t_j}{\longleftrightarrow} v_{t_j} \right\} \quad (2)$$

where $v_1 = v_{t_{1-1}}, v_n = v_{t_{mn}}$.

In VANET, data packets should be forwarded along the routing path in a consecutive way if the former node is connected to the latter node. However, if one node is not in the communication range of the next hop at a certain time stamp, the data packets should be buffered at this node until the next time stamp comes when the consecutive two nodes on the path are connected to each other. Therefore, Equation 2 must fulfill the following requirement.

$$v_{t_k n} = v_{t_{(k+1)1}} = V'_{t_k} \cap V'_{t_{(k+1)}} \quad (k = 1, \dots, m-1) \quad (3)$$

3.3 Connected component

The *connected component* at time stamp t is a non-empty sub-graph of network $G(t)$, in which there exists at least one path for any two vertices. That is, the *connected component* that node v_i is connected with at time stamp t can be denoted as Equation 4.

$$CC(t) = \bigcup_{v_i, v_j \in V_t} \left\{ v_i \overset{t}{\longleftrightarrow} v_j \right\} \quad (4)$$

As mentioned in Section 1, packets are forwarded much faster in a multi-hop way than that in a carry-and-forward style if the source node and the destination node are in the same connected component. Therefore, the performance of the routing strategy can be greatly improved if there are enough stable connected components in VANETs. We will

analyze both the number and the size of the connected components in Sections 4.2 and 4.3.

3.4 Connected component stability

The topology of a VANET keeps on changing due to the movement of the vehicles. For a given connected component $CC(t)$, some vehicles in it might leave while some other new vehicles might come to join at the next consecutive time stamp $(t + 1)$. A good metric to measure the stability of $CC(t)$ should consider both of these two phenomena. For simplification, we use the difference of vertices of the connected component at two consecutive time stamps to measure the stability. Therefore, we can define the stability factor [17] of $CC(t)$ as Equation 5.

$$\zeta_{CC(t)} = \frac{|V_{CC(t)} \cap V_{CC(t+1)}|}{|V_{CC(t)} \cup V_{CC(t+1)}|} \quad (5)$$

It is obvious that $0 \leq \zeta_{CC(t)} \leq 1$. $CC(t)$ is more stable when the value of $\zeta_{CC(t)}$ is larger. The stability of the connected components will be discussed in Section 4.4.

3.5 Location dependency

Besides the topology features of the connected component, the location of the connected component is also important for network design. If the connected component is location dependent, in other words, if the vehicles always form the connected component in a specified region, we do not need to place roadside infrastructures in this region anymore. In order to measure the location dependency, we need the locations of all the vehicles in the connected component. For a given connected component $CC(t)$, let $\Lambda_{CC(t)}$ represent the coordinate set of all the vehicles in $CC(t)$. Then we have Equation 6.

$$\Lambda_{CC(t)} = \{(x_{v_i}, y_{v_i}) \mid v_i \in V_{CC(t)}\} \quad (6)$$

Let $\underline{X} = \min_{v_i \in V_{CC(t)}}(x_{v_i})$, $\underline{Y} = \min_{v_i \in V_{CC(t)}}(y_{v_i})$, $\bar{X} = \max_{v_i \in V_{CC(t)}}(x_{v_i})$, $\bar{Y} = \max_{v_i \in V_{CC(t)}}(y_{v_i})$. The rectangle covering $CC(t)$ can be defined as $\Gamma_{CC(t)} = [(\underline{X}, \underline{Y}), (\bar{X}, \bar{Y})]$, where $(\underline{X}, \underline{Y})$ is the bottom left coordinate of rectangle $\Gamma_{CC(t)}$ and (\bar{X}, \bar{Y}) is the top right coordinate of rectangle $\Gamma_{CC(t)}$. In a consecutive time-stamp set $T_s = \{t_j\}_{j=1}^m$, we denote the region that can cover vehicles in a connected component at *any* time stamp in T_s as $\Psi = \bigcup_{j=1}^m \Gamma_{CC(t_j)}$ and denote the region that can cover vehicles in a connected component at *all* time stamps in T_s as $\Omega = \bigcap_{j=1}^m \Gamma_{CC(t_j)}$. Assume the function $\delta()$ is used to calculate the area of a region, the location-dependency factor of $CC(T_s)$ can be calculated by Equation 7.

$$\xi_{CC(T_s)} = \frac{\delta(\Omega)}{\delta(\Psi)} = \frac{\delta\left(\bigcap_{j=1}^m \Gamma_{CC(t_j)}\right)}{\delta\left(\bigcup_{j=1}^m \Gamma_{CC(t_j)}\right)} \quad (7)$$

It is obvious that $0 \leq \xi_{CC(T_s)} \leq 1$. $CC(T_s)$ is more location dependent when the value of $\xi_{CC(T_s)}$ is larger. The location dependency of the connected components will be discussed in Section 4.6.

4 Dataset analysis and implications

In this section, we will analyze the spatial and temporal features of VANETs by using the concept of connected component other than using the individual node. Based on the discoveries, we will also present our suggestions for network design and optimization.

4.1 Taxi-trace dataset

As is shown in Table 1, we have collected two datasets of real taxi traces. One dataset contains GPS coordinates of more than 533 taxis collected in 20 days in the San Francisco Bay area, USA [18] (SF for short). The other dataset contains GPS coordinates of 13,799 taxis in 9 days in Shenzhen, China (SZ for short). Most of the coordinate-update frequencies of the SF dataset vary from 30 to 60 s, and the SZ dataset has the coordinate-update frequency of about 30 s. Each vehicle has an individual trace file, in which the coordinates together with the time stamps are saved.

To capture the evolving features of network topology, we use linear interpolation to generate consecutive time-synchronized coordinates with a coordinate-update frequency of 30 s. To find the difference between topologies both in on-peak hour and off-peak hour, we select two observation intervals. One is from 1:00 to 2:00 am and the other is from 7:00 to 8:00 am. To explore the variance according to different communication ranges, we choose six communication ranges from 100 to 600 m.

4.2 Size distribution of connected components

Firstly, we take a look at the size distribution of the connected components, which is illustrated by Figures 1, 2, 3 and 4. The x -axis is the fractions of the vehicles in the connected components, which is the size of the connected components divided by the total number of vehicles. The

Table 1 Taxi-trace datasets

	Taxi traces in San Francisco, USA	Taxi traces in Shenzhen, China
Vehicle quantity	533	13,799
Record duration	20 days	9 days
Update frequency	30 to 60 s	30 s
Taxi features	Time stamp, latitude, longitude, occupy status	Time stamp, latitude, longitude, occupy status, velocity, angle
File size	91 MB compressed file	1.06 GB compressed file

Brief summary of taxi traces for analysis.

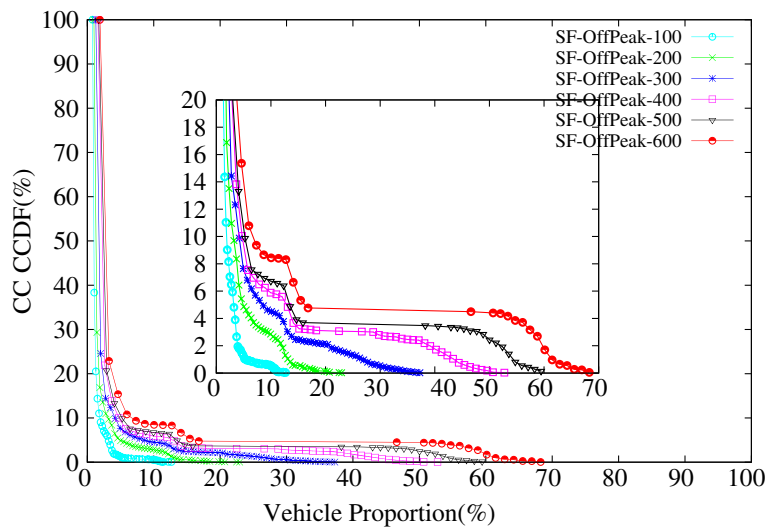


Figure 1 Size distribution of CCs in SF during off-peak hours. The size distribution of connected components in San Francisco, USA during off-peak hours.

y-axis is the complementary cumulative distribution of connected components. From the figures we can learn that most connected components are of very small size. However, the biggest connected component contains a large number of vehicles. A larger communication range can enlarge the size of the connected components and hence slightly increase the fraction of big connected components. Comparing Figure 1 with Figure 2 and Figure 3 with Figure 4, we can conclude that the size distribution of the connected components is independent of the peak hours.

Remark 1. Although the whole topology of the VANET is broken into a large number of small-sized connected components, the biggest connected component can cover a large number of vehicles. As Figure 2 shows, the biggest connected component even contains up to 70% to 80% vehicles. Therefore, if the biggest connected component keeps stable, we can make use of this feature by keeping important information on the vehicles in the biggest connected component and design strategies to forward the information with the biggest connected component.

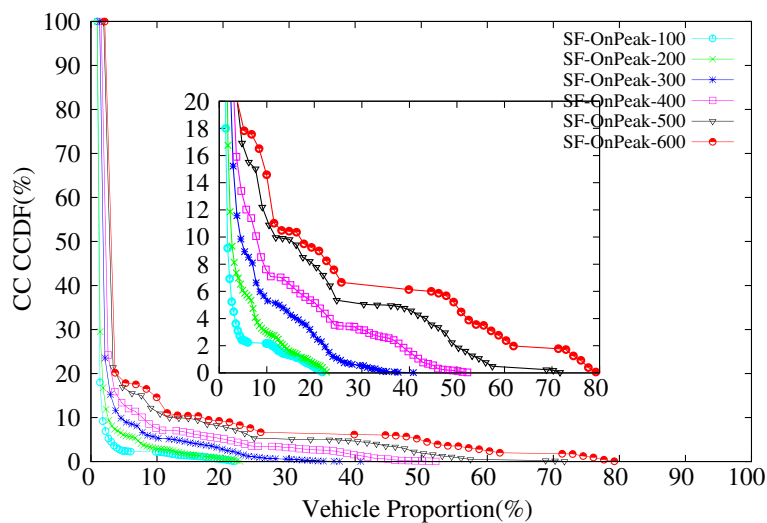


Figure 2 Size distribution of CCs in SF during on-peak hours. The size distribution of connected components in San Francisco, USA during on-peak hours.

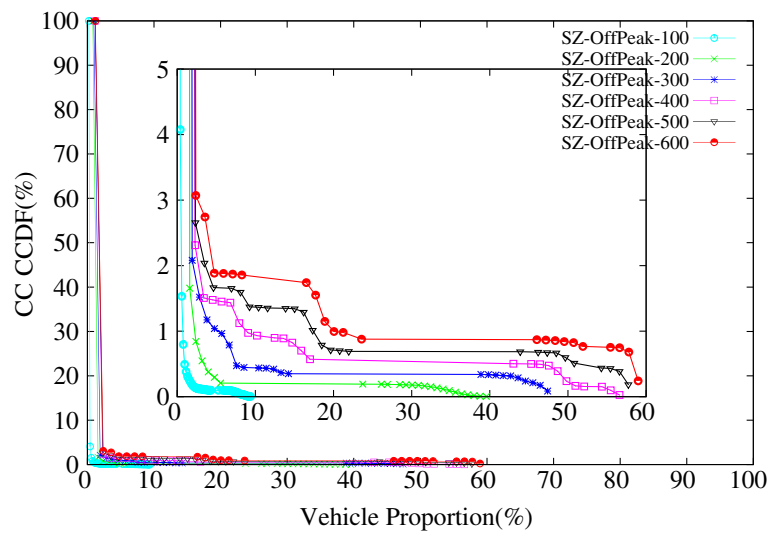


Figure 3 Size distribution of CCs in SZ during off-peak hours. The size distribution of connected components in Shenzhen, China during off-peak hours.

4.3 Size variance of the connected components

Secondly, we measure the size of the biggest connected component for each snapshot of the topology and the results are shown by Figures 5, 6, 7 and 8. The x -axis of each figure is the time stamp with an interval of 30 s. The y -axis of each figure is the number of vehicles in the biggest connected component. As the figures show, the sizes of the biggest connected components are relatively stable. A larger communication range results in a larger size of connected component. Normally there are more vehicles in the peak hours, and the biggest connected components in the peak hour contain more vehicles.

However, there is an exception in Figure 6 after 7:30 am. That is because, for unknown reasons, some of the vehicles did not successfully record their coordinates at a regular frequency (every 30 to 60 s). Therefore, these vehicles are deleted from the results due to the interpolation errors.

Remark 2. It is a piece of good news for routing protocol design that the biggest connected component in each snapshot covers a large number of vehicles. That means many vehicles might benefit from multi-hop forwarding since they get a high probability to be connected

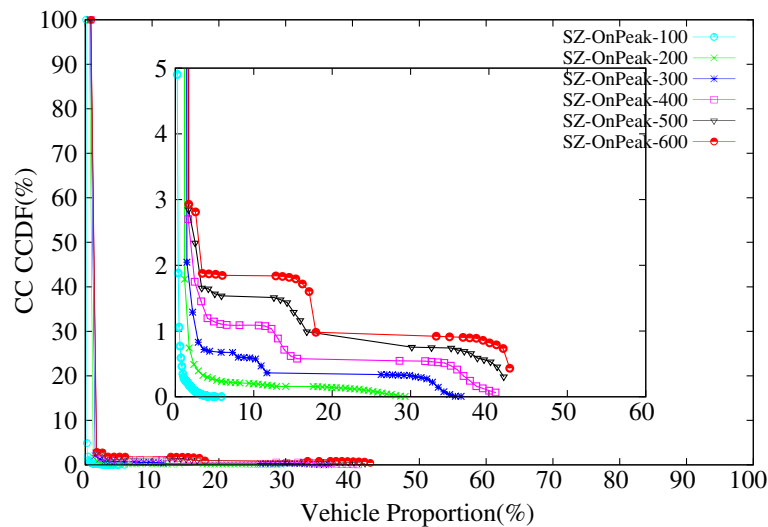


Figure 4 Size distribution of CCs in SZ during on-peak hours. The size distribution of connected components in Shenzhen, China during on-peak hours.

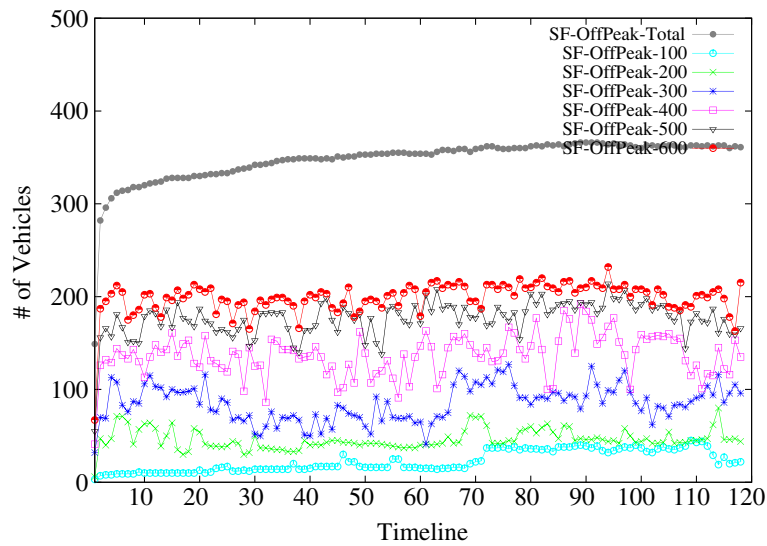


Figure 5 Size variance of biggest CCs in SF during off-peak hours. The size variance of the biggest connected components in San Francisco, USA during off-peak hours.

on the biggest connected component. However, this kind of multi-hop forwarding might be ceased in the next time stamp if the members in the connected component are changed. Thus, we need to further explore the topology evolution of the biggest connected components to verify whether its vehicle members are stable during the dynamic process.

4.4 Stability of connected components

Thirdly, we capture the evolving topology features by the stability factor of every biggest connected component for

each snapshot and plot the results by Figures 9 and 10. Since the interval between two consecutive snapshots is 30 s and the observation duration is 1 h, for each value of the communication range, there are totally 120 values of stability factors. We use the average and the variance of the stability factor to reveal the evolving features of the biggest connected component. In Figure 9, for the San Francisco dataset, there happened to be some vehicles that kept still in the on-peak hours. When communication range is 100, these vehicles take a large proportion in the biggest CC. This results in a large average and a

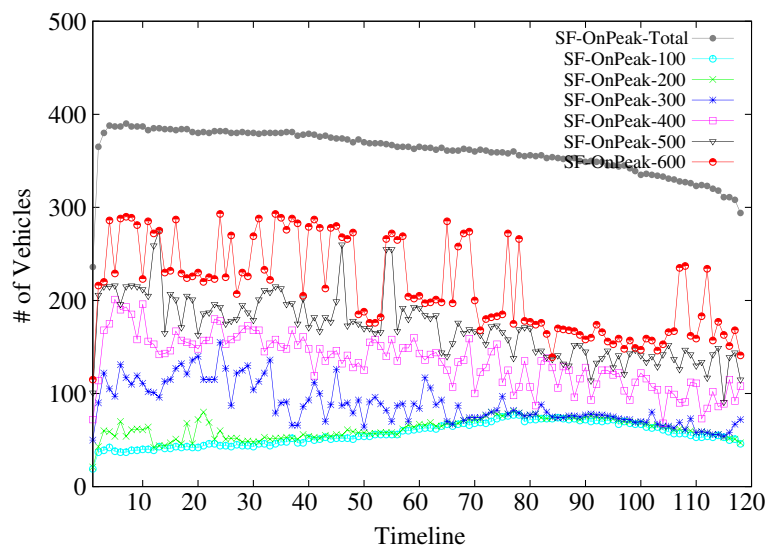


Figure 6 Size variance of biggest CCs in SF during on-peak hours. The size variance of the biggest connected components in San Francisco, USA during on-peak hours.

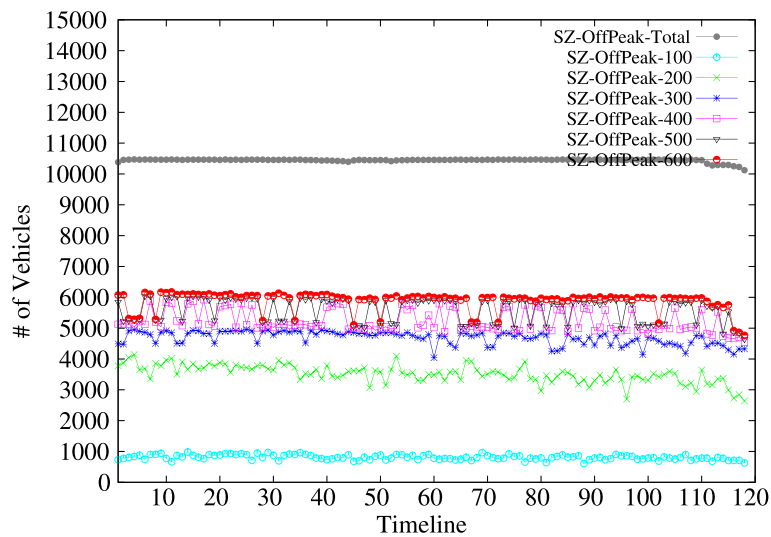


Figure 7 Size variance of biggest CCs in SZ during off-peak hours. The size variance of the biggest connected components in Shenzhen, China during off-peak hours.

small variance of the stability factor for SF-OnPeak. When communication range increases to 200, those unmoving vehicles only take a small proportion in the biggest CC. This results in a decrease of the average and an increase of the variance of the stability factor for SF-OnPeak. In other circumstances, for those vehicles connected to the biggest CC, when the communication range is increasing, they are more likely to keep connected to the biggest CC. Therefore, the average of stability factor gets larger and the variance of stability factor gets smaller. For the same reason, in Figure 10, the average of stability factor

gets larger and the variance of stability factor gets smaller when the communication range is increasing. Results of both Figures 9 and 10 show that the biggest connected component is stable when the communication range is large enough, and the result is not sensitive to on-peak hours or off-peak hours.

Remark 3. When the communication range is larger than 500 m, the averages of the stability factor in both the SF dataset and the SZ dataset are greater than 0.8. This means in average, over 80% of the vehicles in the

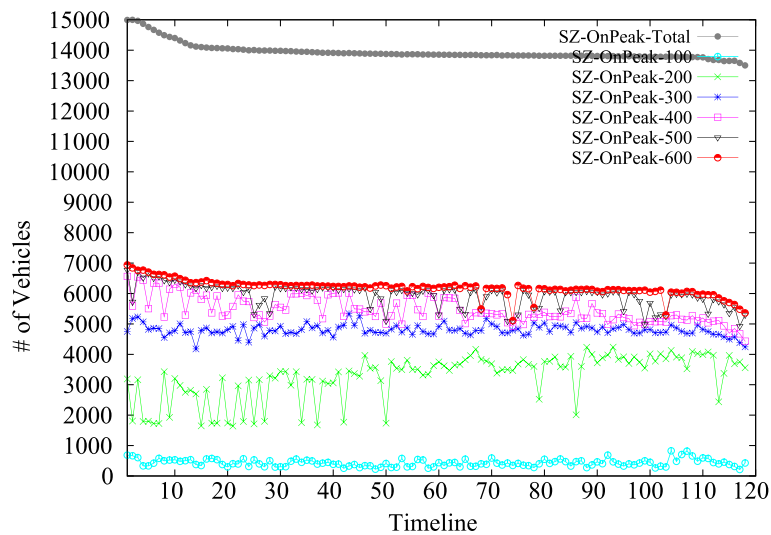


Figure 8 Size variance of biggest CCs in SZ during on-peak hours. The size variance of the biggest connected components in Shenzhen, China during on-peak hours.

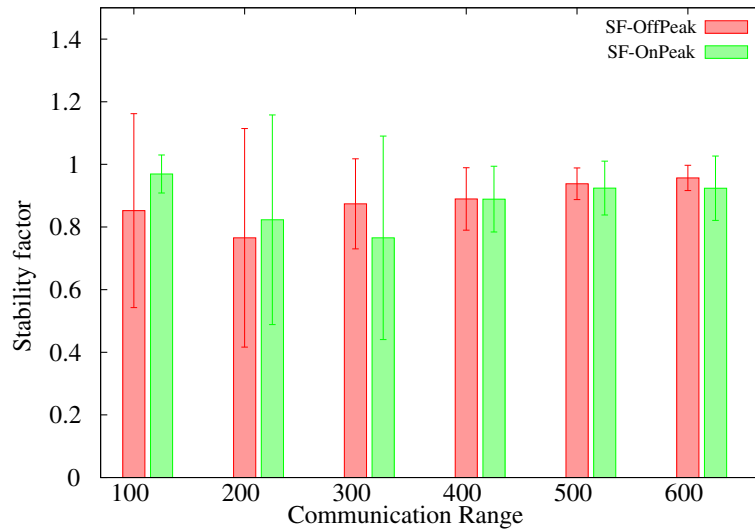


Figure 9 Average and variance of the stability factor in SF. Average and variance of the stability factor of the biggest connected component in San Francisco, USA.

biggest connected component will keep staying in the connected component in the consecutive time interval. If we define those vehicles that always stay in the biggest connected component as *connected component core*, we might improve the performance of the routing strategy if we select the *connected component core* as the packet forwarder, because these *connected component core* vehicles have more opportunities to forward packets by multi-hop routing. However, how to formally define and discover the *connected component core* from spatial and temporal datasets are still open problems.

4.5 Variance of connected components

Section 4.4 reveals that the biggest connected component is relatively stable. However, we need to figure out where the variances take place if the biggest connected component gets changed. Fourthly, we investigate the locations where topology variation takes place. We randomly choose a snapshot of the network topology and plot the coordinates of all the vehicles, of vehicles in the biggest connected component, of vehicles disconnected from the biggest connected component, and of vehicles newly connected with the biggest connected component

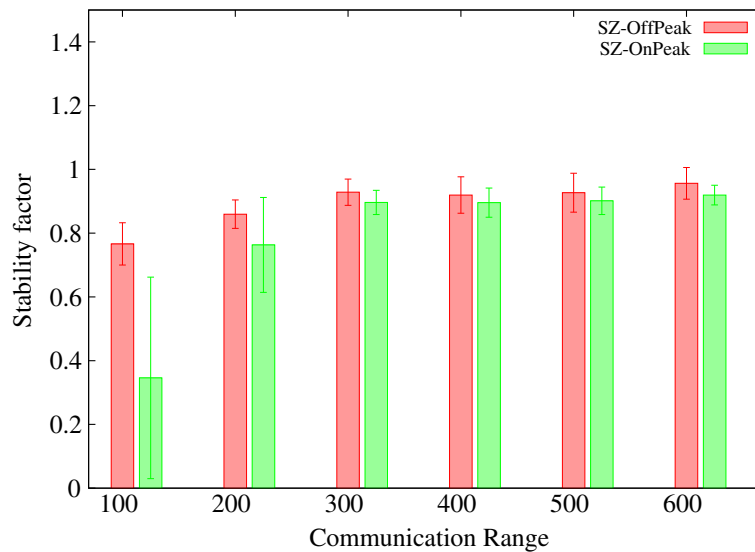


Figure 10 Average and variance of the stability factor in SZ. Average and variance of the stability factor of the biggest connected component in Shenzhen, China.

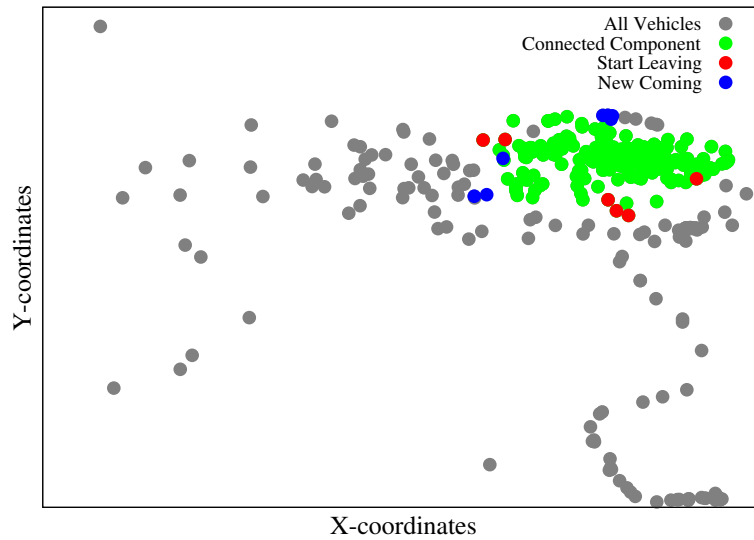


Figure 11 CC variation in SF. Locations for changed vehicles of the biggest connected component in San Francisco, USA.

at this time stamp in Figures 11 and 12. The vehicle distribution shows that variation takes place on those vehicles with low density and on the boundary of the biggest connected component.

Remark 4. It is useful to identify where the variation of the biggest connected component takes place when designing the routing protocol for VANETs. If we define those vehicles where variation takes place as *connected component boundary*, we might also improve the routing performance by not choosing the *connected component boundary* vehicles as forwarders. Again, either defining

or detecting the *connected component boundary* are still unsolved problems.

4.6 Location dependency of connected components

Finally, we study the location dependency of the biggest connected component by calculating the location-dependency factor defined in Section 3.5. The results are given in Figures 13 and 14. It is shown that the communication range is a critical parameter for the location-dependency factor. The biggest connected component only covers a specific region when the communication range is large enough. In the two datasets we use, the

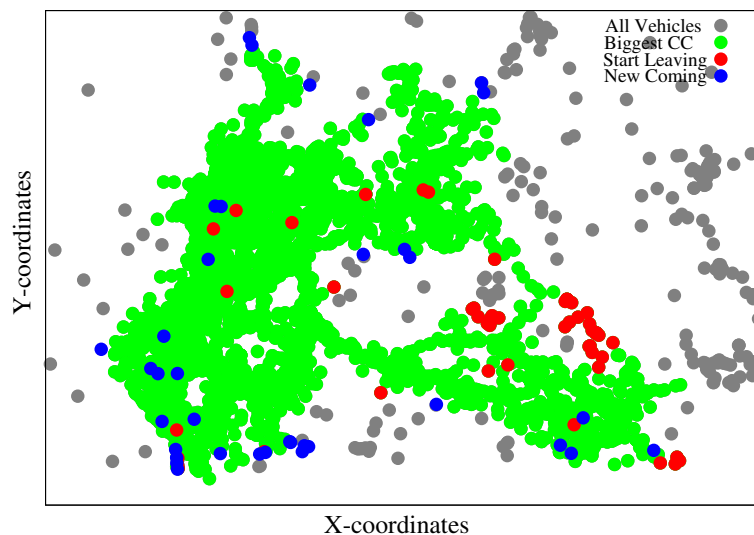


Figure 12 CC variation in SZ. Locations for changed vehicles of the biggest connected component in Shenzhen, China.

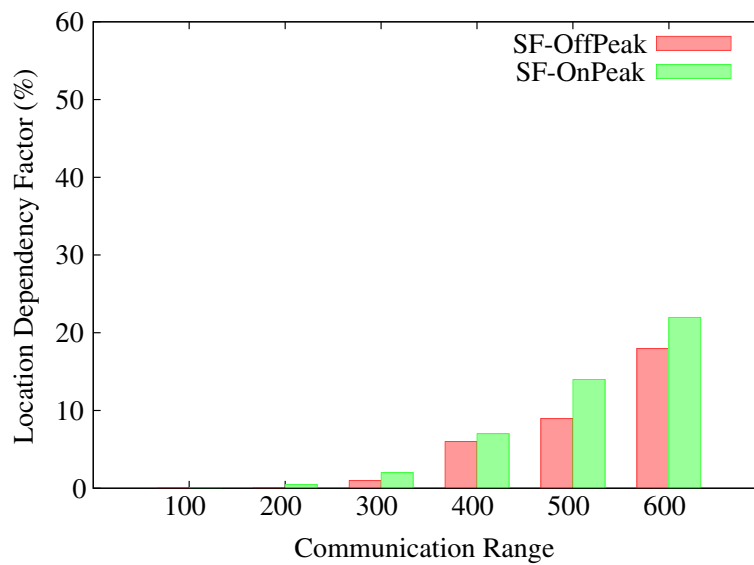


Figure 13 Location-dependency factor of biggest CCs in SF. Location-dependency factor of the biggest connected components in San Francisco, USA.

communication range should be larger than 400 m if we need the location dependency. In more specified cases, when the communication range is 600 m, we can get the results with $\delta(\Psi) \approx 56.24 \text{ km}^2$, $\delta(\Omega) \approx 12.21 \text{ km}^2$ in the SF dataset and $\delta(\Psi) \approx 525 \text{ km}^2$, $\delta(\Omega) \approx 252 \text{ km}^2$ in the SZ dataset. It is reported that the areas of the downtown region in San Francisco and in Shenzhen are 12.25 km^2 [19] and 412 km^2 , respectively. That means that when the communication range is large enough, the biggest

connected component is just located in the downtown region.

Remark 5. The location dependency of the biggest connected component contributes to two aspects in network design. One benefit is that we do not need to deploy roadside infrastructure to the spots where the connected component forms, because the connectivity can be maintained by the connected component. The other benefit

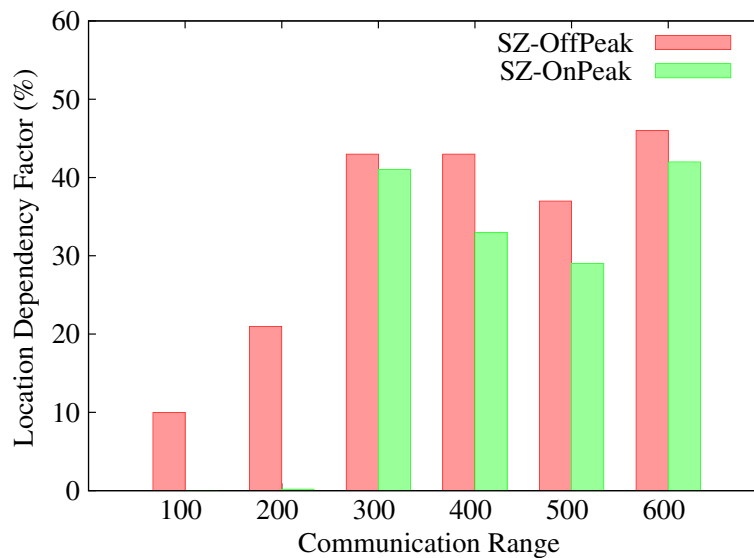


Figure 14 Location-dependency factor of biggest CCs in SZ. Location-dependency factor of the biggest connected components in Shenzhen, China.

is that vehicles in the downtown region should use the multi-hop forwarding strategy rather than the carry-and-forward strategy, since vehicles in this region have a large probability to be connected to the biggest connected component.

5 Conclusions

In this paper, we analyzed the spatial and temporal dynamics of VANETs based on two real taxi-trace datasets. We found that the whole topology of VANETs consists of a large number of small-sized connected components; however, the biggest connected component among them contains a large proportion of vehicles. We also found that the connected component core in the biggest connected component is stable in the case where the communication range is large enough and the variation takes place at the connected component boundary. The performance of the routing protocol might be improved by using the information of the connected component core and connected component boundary. Furthermore, by adopting a reasonably communication range, the biggest connected component has the feature of location dependency, which is very useful to roadside infrastructure deployment and multi-hop packet forwarding. However, how to efficiently find the connected component core and detect the connected component boundary and how to design new network architecture and routing protocols are still open research problems and they are left as our future work.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This paper is partially funded by the National Natural Science Foundation of China under grant No. 61003304, 61105124, 61272485, 61379144 and China Scholarship Council under grant 2011611534, 201206115013. The authors would like to express great appreciation to the reviewers of the paper for their valuable comments on improving the quality of this paper.

Author details

¹College of Computer, National University of Defense Technology, 410073 Changsha, Hunan, China. ²Pillar of Information System Technology and Design, Singapore University of Technology and Design, 138682 Singapore, Singapore. ³College of Information Systems and Management, National University of Defense Technology, 410073 Changsha, Hunan, China.

Received: 1 January 2014 Accepted: 19 October 2014

Published: 3 November 2014

References

1. F Bai, T Elbatt, G Hollan, H Krishnan, V Sadekar, in *Proceedings of IEEE Workshop on Automotive Networking and Applications (AutoNet)*. Towards characterizing and classifying communication-based automotive applications from a wireless networking perspective (San Francisco, CA, USA, 27 Nov–1 Dec, 2006), pp. 1–25
2. V Naumov, TR Gross, in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*. Connectivity-aware routing (car) in vehicular *ad-hoc* networks (Anchorage, Alaska, USA, 6–12 May, 2007), pp. 1919–1927
3. M Abuelela, S Olariu, I Stojmenovic, in *Mobile Ad Hoc and Sensor Systems, 2008. MASS 2008. 5th IEEE International Conference On*. Opera:

- opportunities packet relaying in disconnected vehicular *ad hoc* networks (Atlanta, Georgia, USA, 29 Sept–02 Oct, 2008), pp. 285–294
4. J Jianbin, C Yingwen, X Ming, Y Gu, in *Ubiquitous Intelligence & Computing and 9th International Conference on Autonomic & Trusted Computing (UIC/ATC), 2012 9th International Conference On*. Improving the estimation of residual delay based forwarding method in opportunistic vehicular networks (Fukuoka, Japan, 4–7 Sept, 2012), pp. 79–86
5. J Jeong, S Guo, Y Gu, T He, DH Du, in *Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference On*. Tsf: Trajectory-based statistical forwarding for infrastructure-to-vehicle data delivery in vehicular networks (Genova, Italy, 21–25 June, 2010), pp. 557–566
6. X Cheng, D-Z Du, L Wang, B Xu, Relay sensor placement in wireless sensor networks. *Wireless Network*. **14**(3), 347–355 (2008)
7. X Cheng, X Huang, D Li, W Wu, D Du, A polynomial-time approximation scheme for the minimum-connected dominating set in *ad hoc* wireless networks. *Networks* **42**(4), 202–208 (2003)
8. X Jin, W Su, Y Wei, in *Consumer Communications and Networking Conference (CCNC), 2011 IEEE*. A study of the VANET connectivity by percolation theory (Las Vegas, Nevada, USA, 9–12 Jan, 2011), pp. 85–89
9. D Miorandi, E Altman, Connectivity in one-dimensional *ad hoc* networks: a queueing theoretical approach. *Wireless Network*. **12**(5), 573–587 (2006)
10. Y Zhuang, J Pan, L Cai, in *Proceedings of the Seventh ACM International Workshop on Vehicular InterNetworking*. A probabilistic model for message propagation in two-dimensional vehicular *ad-hoc* networks (Chicago, Illinois, USA, 20–24 Sept, 2010), pp. 31–40
11. A-L Barabasi, Network science. *Philos Trans R Soc*. **371**(1987), 1–3 (2013)
12. R Monteiro, S Sargento, W Viriyasitavat, OK Tonguz, in *Vehicular Networking Conference (VNC), 2012 IEEE*. Improving VANET protocols via network science (Seoul, Korea, 14–16 Nov, 2012), pp. 17–24
13. N Louloulides, G Pallis, MD Dikaiakos, The dynamics of vehicular networks in urban environments (2010). arXiv preprint arXiv:1007.4106
14. D Naboulsi, M Fiore, in *Proceedings of the Fourteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing (ACM MobiHoc)*. On the instantaneous topology of a large-scale urban vehicular network: the Cologne case (Bangalore, India, 29 July–1st August, 2013), pp. 167–176
15. M Piorkowski, N Sarafijanovic-Djukic, M Grossglauser, in *Communication Systems and Networks and Workshops, 2009. COMSNETS 2009. First International*. A parsimonious model of mobile partitioned networks with clustering (Bangalore, India, 5–10 Jan, 2009), pp. 1–10
16. BN Clark, CJ Colbourn, DS Johnson, Unit disk graphs. *Ann. Discrete Math. Bk.* **48**, 165–177 (1991)
17. Y Chen, M Xu, Y Gu, P Li, X Cheng, Understanding topology evolving of VANETs from taxi traces. *Adv. Sci. Technol. Lett.* **42**(Mobile and Wireless), 13–17 (2013)
18. M Piorkowski, N Sarafijanovic-Djukic Grossglauser, CRAWDAD trace set epfl/mobility/cab (v.2009-02-24) (2009). Downloaded from <http://crawdad.cs.dartmouth.edu/epfl/mobility/>
19. MANET driven by SF cab trace. <http://zjije.net/manet/cabtrace.php>

doi:10.1186/1687-1499-2014-180

Cite this article as: Chen et al.: Empirical study on spatial and temporal features for vehicular wireless communications. *EURASIP Journal on Wireless Communications and Networking* 2014 **2014**:180.