

Understanding Topology Evolving of VANETs from Taxi Traces

Yingwen Chen¹, Ming Xu¹, Yu Gu², Pei Li³, Xiuzhen Cheng⁴

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

² Pillar of Information System Technology and Design,
Singapore University of Technology and Design, Singapore

³ College of Information Systems and Management, National University of
Defense Technology, 410073 Changsha, Hunan Province, China

⁴ Department of Computer Science, The George Washington University,
20052 Washington DC, USA

{csywchen@gmail.com, jasongu@sutd.edu.sg, xuming@nudt.edu.cn,
peili@nudt.edu.cn, cheng@gwu.edu}

Abstract. Routing through Connected Components (CC) of VANETs is proved to be much more efficient than carry-and-forward strategies. Understanding the topology evolving features especially those relevant to CCs is very important for routing protocol design and algorithm optimization. This paper explores the network topology and its dynamics based on two real taxi-trace datasets. The size variance of the CCs are analyzed. Considering the time evolution, the stability and variation of the CCs are also examined. Together with the results for topology analysis, implications for routing protocol design and open research issues are then addressed.

Keywords: Connected Components, VANETs, Topology Evolving

1 Introduction

Vehicular Ad hoc Networking, shorten as 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.

Moving vehicles result in a disconnected VANET^[1], thus different kinds of carry-and-forward strategies^{[1][2][3]} have been proposed to support delay tolerant applications. It is also 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^[4]. Therefore, the performance of a VANET's routing strategy can be largely improved if we make full use of the connectivity features of the network.

Since the whole topology of a VANET consists of a large number of **Connected Components (CC)**, we explore the topology and its dynamics to answer questions

such as: whether the CCs have good features for routing protocol optimization; and whether those good features are stable. Based on two real taxi-trace datasets collected from San Francisco, USA, and Shenzhen, China, we analyze the size variance and stability of the CCs. We also investigate the locations of the variation.

2 Dataset analysis and implications

By adopting the UDG model, VANETs topology can be abstracted as follows.

Network. The network $G(T)$ can be expressed by a triple $\langle V, E, T \rangle$, 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 timeline.

Path. There is a non-empty sub-graph of network $G(t)$, denoted as $P = \langle V', E', t \rangle$.

Assume $V' \subseteq V, E' \subseteq E, t \in T, n = |V'|, V' = \{v_{a1}, v_{a2}, \dots, v_{an}\}$. P is called a *Path* iff. there exists $f: V' \rightarrow V', f(v_{ai}) = v_{bi} (i=1, \dots, n)$, s.t. $E' = \bigcup_{j=1}^{n-1} (v_{bj}, v_{b(j+1)})$. Note that v_{b1} and v_{bn} are called the two ends of the path. The path is also denoted as $P = \{v_{b1} | \dots | v_{bn}\}$.

Connected Component. The CC at timestamp t is a non-empty sub-graph of network $G(t)$, in which there exists at least one path for any two vertices. The CC that v_i is

connected with at timestamp t can be formed as $\bigcup_{v_j \in V'} \{v_i | \dots | v_j\}$.

Stability. For a given $CC_i(t)$, some vehicles in it might leave away while some other new vehicles might come to join in at the next consecutive timestamp $(t+1)$. A good metric to measure the stability of $CC_i(t)$ should consider both of these two phenomena. For simplification, we only use the difference of vertices in CCs to measure the

stability. Here we define the stability factor of $CC_i(t)$ as $CC_i(t)_s = \frac{|V_{CC_i(t)} - V_{CC_i(t+1)}|}{|V_{CC_i(t)}|}$.

The performance of the routing strategy can be greatly improved if there are enough stable CCs. This paper aims to understand the basic features of CCs by analyzing the real taxi traces and give some suggestions to routing protocol design.

We have collected two datasets of real taxi traces. One contains GPS coordinates of more than 533 taxis collected in 20 days in San Francisco Bay area, USA^[5] (SF for short). The other contains GPS coordinates of 13,799 taxis in 9 days in Shenzhen, China (SZ for short). Most of the coordinate-update frequencies of SF dataset vary from 30 seconds to 60 seconds, and the SZ dataset has the coordinate-update frequency of about 30 seconds. We use linear interpolation to generate consecutive time synchronized coordinates with coordinate-update frequency of 30 seconds. We select two observation intervals. One is off-peak hour from 1:00 am to 2:00 am, and the other is on-peak hour from 7:00 am to 8:00 am. Six communication ranges vary from 100 meters to 600 meters.

Size variance of the CCs. We measure the size of the biggest CC for each snapshot of the topology and the results are shown by Fig.1. The x-axis of Fig.1 is the timestamp with an interval of 30 seconds. The y-axis of Fig.1 is the number of

vehicles in the biggest CC. As the figures shown, the size of the biggest CCs are relatively stable. A larger communication range results in a larger size of CC.

Remark 1. It is a piece of good news for routing protocol design that the biggest CC 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 on the biggest CC.

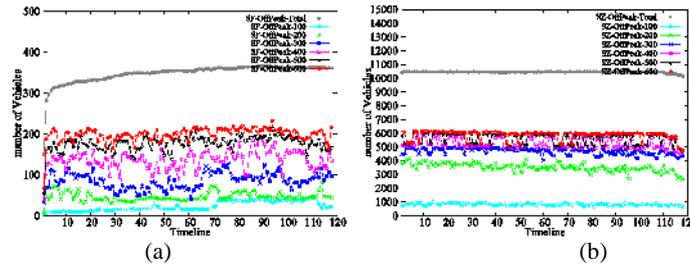


Fig.1. Size variance of the biggest CCs.

Stability of CCs. Secondly, we capture the topology evolving feature with the stability factor of every biggest CC for each snapshot and plot the results by Fig.2. We use the average and the variance of the stability factor to reveal the evolving feature of the biggest CC. Results show that the biggest CC is stable when the communication range is large enough.

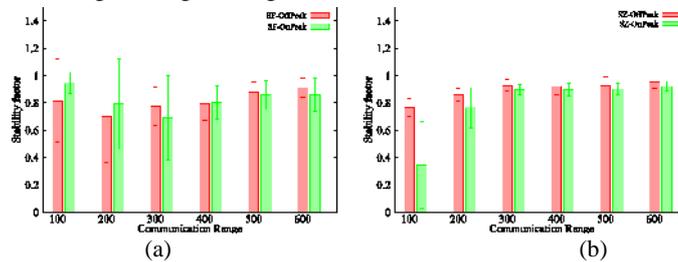


Fig.2. Average and variance of the stability factor.

Remark 2. When the communication range is larger than 500 meters, the averages of the stability factor in both SF dataset and SZ dataset are greater than 0.8. This means in average, over 80% of the vehicles in the biggest CC will keep stay in it. If we define those vehicles that are always stay in the biggest CC as CC core, we might design better routing strategy if we select the CC core as the packets forwarder.

Locations of the variation. Finally, 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 CC, of vehicles disconnected from the biggest CC, and of vehicles newly connected with the biggest CC at this timestamp, in Fig.3. The vehicle distribution shows that variation takes place on those vehicles with low density and on the boundary of the biggest CC.

Remark 3. It is useful to identify where the variation of the biggest CC takes place when designing the routing protocol for VANETs. If we define those vehicles where variation takes place at the *CC boundary*, we might also improve the routing

performance by not choosing the CC boundary nodes as forwarders. However, defining and detecting the CC boundary are still unsolved problems.

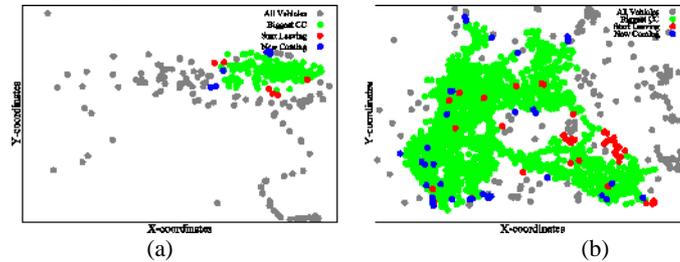


Fig.3. Illustration for the locations where variation takes place.

3 Conclusions

In this paper, we analyzed the topology of VANETs and its dynamics based on two real taxi-trace datasets. We found that the whole topology of VANETs consists of a large number of small-sized CCs, however, the biggest CC among them contains a large proportion of vehicles. We also found the CC core in the biggest CC is stable in case of the communication range is large enough and the variation takes place at the CC boundary. The performance of the routing protocol might be improved by using the information of CC core and CC boundary. But how to efficiently find the CC core and detect the CC boundary are still open research problems and they are left as our future work.

Acknowledgments. This paper is partially funded by National Natural Science Foundation of China under grant No.61003304, 61105124, and China Scholarship Council under grant 2011611534, 201206115013.

References

1. Valery Naumov, Thomas R. Gross.: Connectivity-Aware Routing (CAR) in Vehicular Ad Hoc Networks. In: IEEE 26th International Conference on Computer Communications. IEEE Press, New York (2007)
2. Mahmoud Abuelela, Stephan Olariu, Ivan Stojmenovic.: OPERA: Opportunistic packet relaying in disconnected Vehicular Ad Hoc Networks. In: IEEE 5th International Conference on Mobile Adhoc and Sensor Systems, pp:285-294. (2008)
3. Jianbin Jia, Yingwen Chen, Ming Xu, Gu Yu.: Improving the Estimation of Residual Delay based Forwarding Method in Opportunistic Vehicular Networks. UIC/ATC, pp:79-86. (2012)
4. Aehoon Jeong, Shuo Guo, Yu Gu, Tian He, David HC Du.: Tsf: Trajectory-based statistical forwarding for infrastructure-to-vehicle data delivery in vehicular networks. In: IEEE 30th International Conference on Distributed Computing Systems, pp:557-566. (2010)

5. Michal Piorkowski, Natasa Sarafijanovic-Djukic, Matthias Grossglauser. {CRAWDAD} tracesetepfl/mobility/cab (v. 2009-02-24). Downloaded from <http://crawdad.cs.dartmouth.edu/epfl/mobility/cab> (2009)