DIAL: A Distributed Adaptive-Learning Routing Method in VDTNs

Bo Wu, Haiying Shen and Kang Chen
Department of Electrical and Computer Engineering
Clemson University, Clemson, South Carolina 29634
{bwu2, shenh, kangc}@clemson.edu

***Abstract*—In this paper, we focus on designing an efficient routing method in Vehicle Delay Tolerant Networks (VDTNs) without centralized infrastructures. Current routing methods either have limited routing efficiency or need the help of centralized infrastructures to improve the routing performance, which is deviated from the goal of building VDTNs without centralized infrastructures. In this paper, we propose a DIstributed Adaptive-Learning routing method for VDTNs, namely DIAL, by taking advantage of the human beings' communication feature, that is, most interactions are generated by pairs of people who interacted often previously. DIAL consists of two components: the information fusion based routing method and the adaptive-learning framework. The information fusion based routing method enables DIAL to improve the routing performance by sharing and fusing multiple kinds of routing information without centralized infrastructures. Furthermore, based on the information shared by information fusion based routing method, the adaptive-learning framework enables DIAL to design personalized routing strategies for different vehicle pairs without centralized infrastructures. Therefore, DIAL can not only share and fuse multiple kinds of routing information of each vehicle without centralized infrastructures, but also design a personalized routing strategy for each vehicle pair. Extensive trace-driven simulation demonstrates that DIAL has better routing success rate and shorter average delay in comparison with state-of-the-art routing methods, which need the help of centralized infrastructures in VDTNs.**

I. INTRODUCTION

Vehicle Delay Tolerant Networks (VDTNs) is a new Delay Tolerant Networks (DTNs) concept built on the top of VNETs (Vehicle NETworks) where vehicles act as the communication infrastructures. Currently, the performance of routing on VDTNs [1, 2] can not satisfy the need of real applications due to features of VDTNs such as high vehicle mobility, dynamic scenarios, sparsity of vehicles, short contact durations, disruption and intermittent connectivity. In order to improve the routing performance, preprocessed information and centralized infrastructures are used by some routing methods [3–7]. However, such kind of routing methods are deviated from the goal of building VDTNs without centralized infrastructures. In this paper, we propose a DIstributed Adaptive-Learning routing method, namely DIAL, which can improve the VDTNs' routing efficiency without centralized infrastructures.

The current DTN routing methods can be divided to three categories by the information they used: contact based routing [1], centrality based routing [2, 3] and location based routing [4–7] methods. They all have certain drawbacks as follows. The contact based routing methods [1, 8] deliver packets gradually by relaying vehicles with higher probability to reach the target vehicles. However, VDTNs usually consist of thousands of sparsely distributed vehicles (nodes), leading to a low chance for a packet to encounter a suitable relay vehicle. Centrality based routing methods [2, 9] improve the routing efficiency by exploiting different centrality criteria such as degree and betweenness in which the multi-hop information can be considered. However, the reachability of vehicles to different vehicles is not considered and the centrality criteria only measure the importance of vehicles globally. Also, all the packets are tending to be forwarded by very few vehicles with very high centrality, which may lead to a load imbalanced problem. The location based routing methods [4–7] usually take advantage of the extra location information we can get such as GPS navigation trajectories and frequently visited locations, which can significantly improve the routing efficiency. But there are still two problems. First, not every vehicle has a GPS navigation system. Second, even if each vehicle has such information, we still need the help of centralized infrastructures to share the information among different vehicles.

To deal with these problems, we compare the performances of different routing methods in the analysis. We find two useful observations as follows: (i) The performances of different routing methods can be different on different vehicle pairs. (ii) It is true that there are some correlations between the routing performances of different methods and the features of vehicle pairs (e.g. geographic distance of the two vehicles and whether two vehicles are in the same community). For example, if two vehicles are in the same community, contact based routing can be more efficient than centrality and location based routing methods. However, when it comes to a specific vehicle pair, such correlations can be influenced by many unknown factors, which makes it difficult to predict the performances of different routing methods. Based on observation (i), we hope to choose the routing method for each vehicle pair separately so that the routing performances on all the vehicle pairs can be optimal. However, based on observation (ii), we find that it's difficult to choose the routing methods when it comes to a specific vehicle pair. Even if we can choose the routing methods based on the features of vehicle pairs, centralized infrastructures are needed for sharing the necessary information among vehicles in order to calculate those features such as geographic distances and whether two

978-1-4673-9948-7/16 $31.00 © 2016 IEEE
DOI 10.1109/IoTDI.2015.12

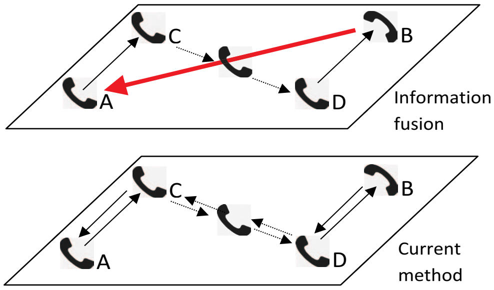


Fig. 1: Current methods vs. information fusion based routing method.

vehicles are in the same community. Fortunately, the routing of VDTNs is different from the general DTNs. General DTNs can consist of any kinds of moving objects, but VDTNs consist of vehicles which are driven by human beings. Hence, the communication on the top of VDTNs is actually the communication between human beings. Previous studies [10] show that most interactions are generated by pairs of people who interacted often previously. By taking advantage of this feature, we design a DIstributed Adaptive-Learning (DIAL) routing method which can not only share the important routing information of each vehicle without centralized infrastructures, but also design a personalized routing strategy for each vehicle pair. DIAL consists of two components: information fusion based routing method and adaptive-learning framework.

By taking advantage of the human beings communication feature we mentioned above, the information fusion based routing method enables DIAL to improve the routing performance by sharing and fusing multiple kinds of routing information without centralized infrastructures. In the information fusion based routing method, vehicle A attaches its personal information of frequently visited locations and frequently contact vehicles to the packet when vehicle A tries to deliver a packet to vehicle B. Then, based on centrality based method, the packet is delivered one relay vehicle by another relay vehicle. Once vehicle B receives all the personal information of vehicle A, vehicle B fuses all the personal information by setting different thresholds for adopting different routing methods. Next time, vehicle B delivers the packet to vehicle A by choosing more efficient routing method. Since human beings tend to interact with others who they interacted with before, the information fusion based routing method avoids the inefficient routing in the subsequent communication between A and B and hence improves the routing performance without any centralized infrastructures. Figure 1 shows the difference between current routing method and the information fusion based routing method. In a current routing method, vehicle B uses a determined routing method to to deliver a packet to vehicle A. In the information fusion based routing method, vehicle B chooses the most efficient routing method among multiple methods based on the personal information sent by vehicle A to deliver a packet to vehicle A.

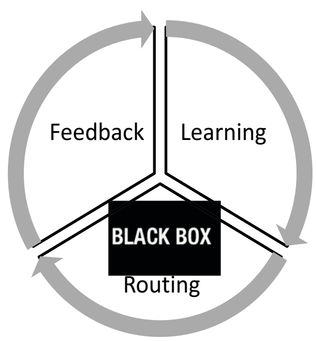


Fig. 2: The adaptive-learning framework

With the help of the information fusion based routing method, we can share and fuse information of vehicles without centralized infrastructures. However, based on the observation (ii), we find that it's difficult to predict which routing method should be the best option for different vehicle pairs even though there are some correlations between the routing performances of different methods and the features of the vehicle pairs. Therefore, in order to design a personalized routing strategy for each vehicle pair, we build an adaptive-learning framework. Instead of determining the thresholds for different methods statically and globally, in the adaptive-learning framework, we consider the routing process as a black box and use the feedback of success rates to determine the thresholds dynamically as shown in Figure 2. Similar as the information fusion based routing method, by taking advantage of the human beings' communication feature, we can calculate the routing success rates of different routing methods which use different information for a given pair of vehicles. Then based on the feedback of the success rates, the vehicle pair can analyze the performances of different routing methods and adjust the routing strategies. For example, vehicle B frequently receives the packets sent from vehicle A. These packets can be delivered by contact based method, centrality based method or location based method. When vehicle B sends packets to vehicle A, vehicle B sends the numbers of packets successfully delivered by different methods from vehicle A at the same time. Then vehicle A can calculate the success rates based on the numbers of packets successfully delivered by different methods and adjust thresholds for different routing methods accordingly to give preference to the method that can lead to the highest success rate. The routing strategy can be self-adaptive in the adaptive-learning framework as shown in Figure 2. Therefore, different vehicle pairs may use different routing strategies and at the same time, the routing strategies can be continually improved according to the feedbacks of the routing performances.

The main contributions of this paper are as follows:

1) We improve the routing efficiency of VDTNs by taking advantage of two features: (i) Most interactions are generated by pairs of people who interacted often previously; (ii) The performances of different routing methods can be different on different vehicle pairs.
2) We design the information fusion based routing method. This method can distributedly share and fuse the vehicles' personal information in the routing process without the help of centralized infrastructures. Therefore, it is more practical and efficient than previous routing methods which need the help of centralized infrastructures.
3) We design an adaptive-learning framework on the top of the information fusion based routing method, which can

design different routing strategies for different vehicle pairs for more efficient VDTN routing than the basic information fusion based routing method.

The rest of this paper is organized as follows. Section II presents the related work. Section III explains the rationale of DIAL by measuring the routing performances of different methods, analyzing the reason and discussing the solution. Section IV introduces the detailed design of DIAL. In Section V, the performance of DIAL is evaluated by trace-driven experiments. Section VI summarizes the paper with remarks on our future work.

II. RELATED WORK

Based on the methods to access the routing information, the current routing methods can be divided to two categories: distributed routing method (which accesses the routing information distributedly) and centralized routing method (which accesses the routing information with the help of centralized infrastructures). In the category of distributed routing method, PROPHET [1] simply selects vehicles with higher encounter frequency with target vehicles for relaying packets. PeopleRank [2] is inspired by the PageRank algorithm, which calculates the rank of vehicles and forwards packets to the vehicles with higher ranks. In the category of centralized routing method, SimBet [3] identifies some bridge nodes as relay nodes which can better connect the VNETs by centrality characteristics to relay packets. AAR [4] defines the frequently visited locations of each vehicle and tries to deliver the packets to the frequently visited locations of target vehicles. GeOpps [5] directly obtains the future location of the target vehicle from GPS data and spreads packets to certain geographical locations for routing opportunities through the shortest paths. GeoDTN [6] encodes historical geographical movement information in a vector to predict the possibility that two vehicles become neighbors. DTN-FLOW [7] divides the map to different areas and predicts the future visiting area of vehicles, which improves the routing performance since it is much easier to predict the future visiting areas than exact future locations.

Based on the type of information applied, the above routing methods can be divided to three categories: contact based routing method, centrality based routing method and location based routing method as we mentioned in Section I. In the above introduced methods, PROPHET [1] belongs to the contact based routing method; PeopleRank [2] and SimBet [3] belong to the centrality based routing method; AAR [4], GeOpps [5], GeoDTN [6] and DTN-FLOW [7] belong to the location based routing method. These methods have certain drawbacks as indicated in Section I.

III. RATIONALE

There are many works comparing the overall performances of different routing methods. However, there lacks a comprehensive analysis when it comes to the performances of different vehicle pairs with different features. Therefore, in this section, we measure the success rates of different routing methods on vehicle pairs with different features using two real world VNET traces gathered by taxi GPS in different cities, referred to as *Roma* [11] and *SanF* [12]. The *Roma* trace contains mobility trajectories of 320 taxies in the center of Roma from Feb. 1 to Mar. 2, 2014. The *SanF* trace contains mobility trajectories of approximately 500 taxies collected over 30 days in San Francisco Bay Area. For the routing methods, as introduced in Section II, we choose AAR [4], Prophet [1] and PeopleRank [2] which represent location, contact and centrality based routing methods, respectively. For the features of vehicle pairs, we measure the contact distance, the geographic distance and the centrality of a vehicle pair which are defined as follows.

1) **Contact distance of a vehicle pair:** We first transfer each trace to a *contact graph* based on the contact duration. The nodes of the graphs are the taxies in the traces, the edges are the contacts between pairs of taxies. We naturally think that if two vehicles encounter each other more often, they are in a closer relationship and only the contacts which have accumulative durations long enough can be considered as edges. To be more specific, we define an accumulative contact duration threshold (3000s in *Roma* and 5000s in *SanF*) and the contacts with accumulative durations larger than the threshold can be considered as edges. In the contact graphs, we calculate the contact distance of two vehicles as the number of hops in the shortest path between the two vehicles.
2) **Geographic distance of a vehicle pair:** AAR finds the active area for each vehicle, where it frequently visits. Here, we define the geographic distance of a vehicle pair as the shortest distance between the two vehicles' active areas.
3) **Centrality of a vehicle pair:** We define the centrality of a vehicle as the PageRank value of the vehicle. The centrality of a vehicle pair is the sum of the two vehicles' centralities.

A. Measurement Study

We run contact, centrality and location based methods based on the traces. Firstly, we randomly pick 1000 vehicle pairs (500 pairs in *Roma* and 500 apirs in *SanF*) and use contact, centrality and location based routing methods to deliver a packet between each vehicle pair simultaneously. Then, we analyze the performance of different routing methods on different vehicle pairs. Figure 3 shows the percentage of vehicle pairs that each routing method performs the best. The experimental result follows Location>Centrality>Contact. However, when it comes to individual vehicle pairs, the location-based routing method performs the best on 47% of the vehicle pairs. The centrality based routing method performs the best on 42% of the vehicle pairs and the contact based routing method performs the best on 11% of the vehicle pairs. Therefore, we cannot conclude that one routing method is better than another routing method for every vehicle pair although the overall success rates are comparable. Actually, although location

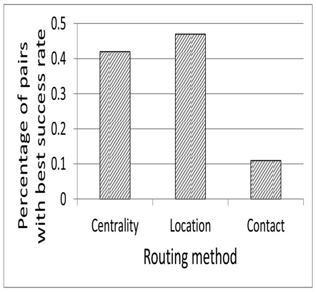


Fig. 3: The diversity of different routing methods.

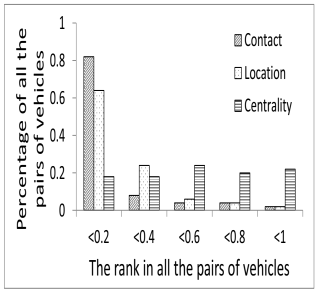


Fig. 4: The top 50 vehicle pairs with the shortest delays of contact routing method.

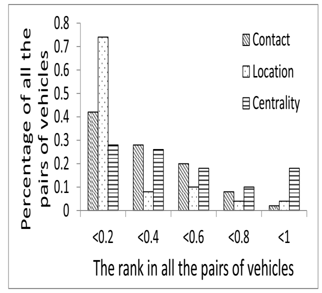


Fig. 5: The top 50 vehicle pairs with the shortest delays of location routing method.

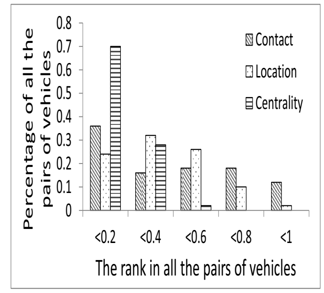


Fig. 6: The top 50 vehicle pairs with the shortest delays of centrality routing method.

routing method which has the highest success rate can perform the best on a lot of vehicle pairs, contact routing method which has the lowest success rate can still perform the best on some vehicle pairs.

Next, we try to figure out the reason why different routing methods have different routing performances on different vehicle pairs. Firstly, in order to find the reason why location based routing method performs better than centrality and contact based routing methods, we select top 50 vehicle pairs with the shortest delays when using contact routing method. Then, we compare their contact distances with geographic distances and centralities. Since contact distances, geographic distances and centralities cannot be compared directly, we transfer the values of different metrics to the ranks in all the vehicle pairs in the analysis. For example, for the contact distance, if the contact distance is 0.25, it means that the contact distance is longer than 25% of all the vehicle pairs; for the geographic distance, if the geographic distance is 0.25, it means that the geographic distance is longer than 25% of all the vehicle pairs; for the centrality, if the centrality is 0.25, it means that the centrality is smaller than 25% of all the vehicle pairs. Figure 4 compares the contact distances with geographic distances and centralities of vehicle pairs. As shown in Figure 4, those vehicle pairs tend to have a relatively closer contact distances in all the vehicle pairs and at the same time have relative longer geographic distances and smaller centralities in all the vehicle pairs. This observation is reasonable and accounts for the phenomenon that these vehicle pairs achieve the shortest delays when they use the contact routing method.

Sometimes, even though the contact metric is not as good as the centrality and location metrics, the contact routing method still performs the best. For example, as shown in Figure 7, suppose two vehicles A and B are in the same community of the Electrical and Computer Engineering Department of Clemson University. However, the vehicles in the ECE community are distributed in two different locations: Clemson University International Center for Automotive Research (CU-ICAR) in Greenville and Clemson. Vehicle A is in Clemson and vehicle B is in Greenville. In this scenario, we should put a particular emphasis on the contact based routing method if A wants to send B a packet. Suppose A has a location utility of 50 miles (the vehicle's frequently visited location is 50 miles away from Clemson), a centrality utility of 100 (the vehicle can meet 100 cars a day) and a contact utility of 0.09 (the vehicle meet B with probability 0.09). Now vehicle A meets three vehicles: a vehicle in Greenville with much higher location utility 10 miles (the vehicle's frequently visit location is 10 miles away from Clemson), a vehicle with a much higher centrality utility 1000 (the vehicle can meet 1000 cars a day) and a vehicle with a little higher contact utility 0.1 (the vehicle meet B with probability 0.1). If A chooses the location based routing method, it chooses the vehicle in Greenville with a location utility 10 miles. Then, the packet may approach Greenville faster by current relay vehicle. However, it cannot be guaranteed that the packet can still easily find the next vehicle which is going to ICAR in Greenville since most of the vehicles in Greenville are not going to ICAR. Therefore, the location based method may cause a failure although we think we select a suitable vehicle with a little high location utility at one of the hops. Also, if A chooses vehicle which is very active as the relay vehicle, the relay vehicle may visit a lot of places but it is very likely that the relay vehicle won't visit Clemson at all since the map is very large. Therefore, the centrality based method may also cause a failure. On the contrary, A may just choose a vehicle with a higher contact utility than itself. But since A and B are in the same community, it is guaranteed that there must be better choices one after another and it is with high probability that contact based method can successfully deliver the packet at last. Therefore, in this scenario, maybe it's a better choice to choose the contact method and a relay vehicle with a higher contact utility rather than the location and centrality methods.

Similarly, we select top 50 vehicle pairs with the shortest delays when using location routing method. Then, we compare their geographic distances with contact distances and centralities. Figure 5 compares the contact distances with geographic distances and centralities of these vehicle pairs. As shown in Figure 5, these vehicle pairs tend to have a relatively closer geographic distances in all the vehicle pairs and at the same time have relative longer contact distances and smaller centralities in all the vehicle pairs. From Figure 4 and Figure 5,

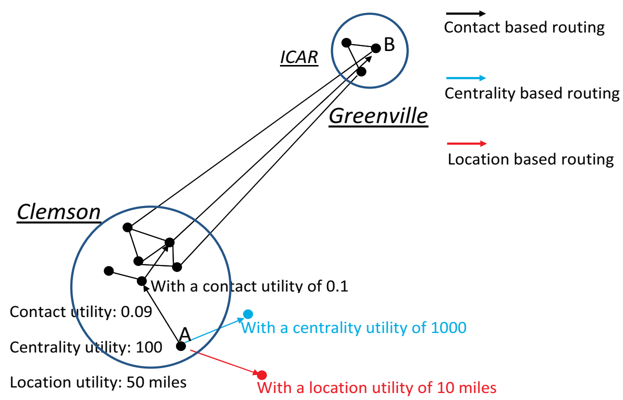


Fig. 7: An example which shows the correlation between routing performances of different methods and the features of the vehicle pairs.??change fig

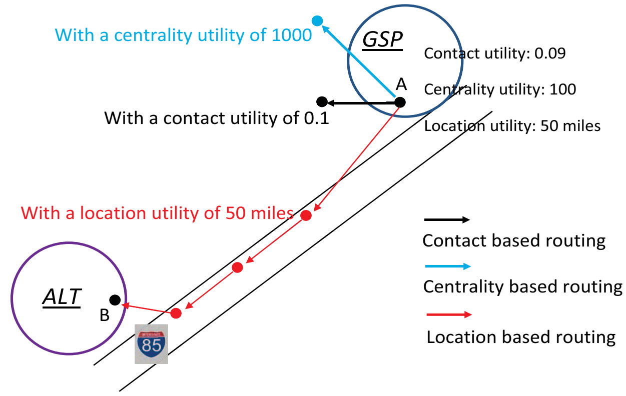


Fig. 8: An example which shows some random factors in the routing.

we can also see that vehicle pairs which have relatively high contact (geographic) distances tend to have relatively high geographic (contact) distances too. The reason may be that vehicles which have closer frequently visited locations tend to meet each other with higher possibilities. However, as shown in Figure 4 and Figure 5, a relatively high contact (geographic) distances cannot guarantee relatively high geographic (contact) distances since there are many other factors that influence vehicles' contact distances. Finally, we select top 50 vehicle pairs with the shortest delays of centrality routing method. Then, we compare their geographic distances with contact distances and centralities in Figure 6. As shown in Figure 6, these vehicle pairs tend to have relatively higher centralities in all the vehicle pairs and at the same time have relatively longer contact and geographic distances in all the vehicle pairs.

B. Analysis

Based on the above analysis, we find that there are some correlations between the features (contact distance, geographic distance and centrality) of vehicle pairs with the performances of different routing methods on them. However, as we can see from Figure 4, Figure 5 and Figure 6, such correlations are not always very clear. For example, as shown in Figure 5, there are still many vehicle pairs in which location routing method performs well but at the same time, with relatively long geographic distances. We further analyze the reason and find that geographic distance is not the only factor that influences the performance of location based routing method. For example as shown in Figure 8, suppose that two vehicles A and B are in two communities which named GSP airport and ALT airport, respectively. Also, vehicles A and B are with a long geographic distance of 100 miles. In this scenario, we should put a particular emphasis on the location based method if A wants to send B a packet even though A and B are with a long geographic distance. Suppose A has a location utility of 100 miles (vehicle A's frequently visited location is 100 miles away from vehicle B's frequently visited location), a centrality utility of 100 (the vehicle can meet 100 cars a day) and a contact utility of 0.1 (the vehicle meets B with probability 0.1). Suppose vehicle A meets three vehicles: a vehicle with a little higher location utility 50 miles (the vehicle's frequently visited location is 50 miles away from vehicle B's frequently visited location), a vehicle with a much higher centrality utility 1000 (the vehicle can meet 1000 cars a day) and a vehicle with a much higher contact utility 0.5 (the vehicle meets B with probability 0.5). If A chooses contact based routing method, it chooses the vehicle with much higher contact utility 0.5. Then, the packet may approach B on the contact graph faster by current relay vehicle. However, it cannot be guaranteed that the packet can still easily find the next vehicle with higher contact utility since we cannot ensure that the packet can reach B's community. Therefore, the contact based method may cause a failure although vehicle A selects a suitable vehicle with a little high contact utility in one of the hops. Also, if A chooses vehicle which is very active as the relay vehicle, the relay vehicle may visit a lot of places but it is very likely that the relay vehicle won't visit ALT airport at all since the map is very large. Therefore, the centrality based method may also cause a failure. On the contrary, A may just choose a vehicle with a little higher location utility (50 miles) than itself even if 50 miles seems still far away from B. However, we can see from Figure 8 that both of their frequently visited locations are on road 85 although two locations are far from each other. Road 85 has many vehicles and hence many relay vehicles. Therefore, we can make sure there will be more suitable location based relay vehicles for the next hops. Therefore, we would like to choose location based method and a vehicle with only a little higher location utility.

C. Challenge and Solution

Based on the above analysis, we find that vehicle pairs usually have their unique situation which may be very complex and it is necessary to design a unique routing method for each vehicle pair in VDTNs in order to take advantage of different routing methods simultaneously. However, when it comes to real implementation, it is challenging to find the best routing method for each vehicle pair for the following reason. There are many factors which can influence the performances of different routing methods. It is impossible for us to design corresponding strategy for each of them. Furthermore, even if we can list as many as factors and design corresponding strategy for each of them, it still will not be the best way since the situation for each vehicle pair is a combination of different factors.

In order to solve the problem, first of all, we give a threshold to the corresponding utility of each routing method. In addition, to continually send packets to relay vehicles with higher the same kind of utility, the first relay vehicle (so as the following relay vehicles) must have a utility larger than the corresponding threshold. A higher threshold means the corresponding routing method is less suitable for the vehicle pair and we won't choose it unless the current encountered vehicle has a relatively high routing utility (i.e., has very high chance to deliver the packet). A lower threshold means the corresponding routing method is more suitable for the vehicle pair. Therefore, although the current encountered vehicle may not have very high chance to deliver the packet, there must be more chances in the future.

Then, we ignore the detailed factors in the micro-scope and only focus on the routing success rates of different routing methods. We believe that a higher success rate can reflect the underestimate of the corresponding method in the system and hence we can decrease the threshold of the method. On the contrary, we can also increase the threshold of the method. In order to calculate the threshold, we let the target vehicle record the numbers of successfully delivered copies delivered by different routing methods and let the source vehicle record the numbers of copies delivered by different routing methods. Once the target vehicle receives the copies, it sends a packet to the source vehicle with the information of successfully delivered copies delivered by different routing methods. In this way, we can jump over the difficulty of analyzing different factors from a micro-scope and at the same time, each vehicle pair can learn its own optimized thresholds continually.

IV. SYSTEM DESIGN

Before introducing the detailed design of DIAL, we first give an overview of DIAL. DIAL consists of two components: information fusion based routing method and adaptive-learning framework. As we introduced in Section II, in order to improve the routing performance, current routing methods take advantage of the information shared by centralized infrastructures, which is deviated from the initial goal of building DTNs. By taking advantage of the human beings' communication feature mentioned above, the information fusion based routing method enables DIAL to improve the routing performance by sharing and fusing multiple kinds of routing information without centralized infrastructures. First, two vehicles adopt centrality based method to achieve the first communication. In the first communication, two vehicles store the frequently visited locations and frequently encountered vehicles of each other. Then, when the two vehicles need to communicate again, based on more detailed information of target vehicle which includes the frequently visited locations and frequently encountered vehicle, they choose the best routing method for communication. At the same time, the information of the frequently visited locations and frequently encountered vehicles of target vehicle is updated from time to time during the communication. Therefore, the information fusion based routing method can share multiple kinds of routing information of vehicles in the network and choose different routing method to deliver packets based on the different information. At the same time, in order to balance the numbers of copies of a packet sent by different routing methods and optimize the routing performance, we set each routing method with a threshold of utility based on the overall routing efficiency of the method. In each routing method, the selected relay vehicles not only need to have a higher utility of the corresponding method than the previous relay vehicles, but also need to have a higher utility than the corresponding threshold of the method.

In information fusion based routing method, the thresholds for different methods are static. However, from observation (i), the performances of different methods can be different on different vehicle pairs. Therefore, we design an adaptive-learning framework which further enables DIAL to design personalized routing strategies for different vehicle pairs without centralized infrastructures. Similar as the information fusion based routing method, by taking advantage of the human beings' communication feature, we can calculate the routing success rates of different routing methods which use different information. Then, based on the feedback of the success rates, we can analyze the performances of different routing methods and adjust the routing strategies. For example, vehicle B frequently receives the packets sent from vehicle A. These packets can be delivered by contact based method, centrality based method or location based method. When vehicle B sends packets to vehicle A, vehicle B sends the numbers of packets successfully delivered by different methods from vehicle A in the last time. Then, vehicle A can calculate the success rates based on the numbers of packets successfully delivered by different methods and adjust thresholds for different routing methods accordingly to give preference to the method that can lead to the highest success rate. The routing strategy can be self-adaptive in the adaptive-learning framework as shown in Figure 2. Therefore, we can determine different routing strategies for different vehicle pairs and at the same time the routing strategies can be adaptively changed according to the feedbacks of the routing performances.

In the following part of this section, we introduce the details of the information fusion based routing method and adaptive-learning framework, respectively.

A. Information Fusion based Routing Method

In the information fusion based routing method, we first introduce the initial routing method of delivering a packet from vehicle A to vehicle B when vehicle A and vehicle B never communicated before. Then, we introduce a data structure on vehicle B named *address book* which stores the frequently visited locations and frequently encountered vehicles of vehi-

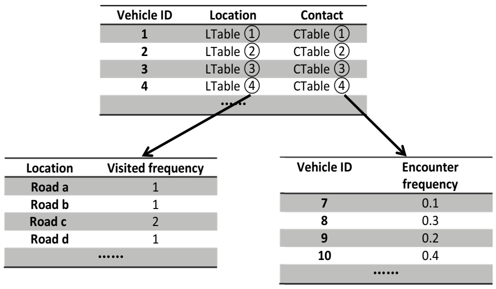


Fig. 9: An example of address book

cles which send packets to vehicle B frequently. Finally, we describe the process of routing process from B to A based on the personalized information of vehicle A stored in the address book of B.

1) Initial routing method: We adopt PeopleRank [2], which is a centrality based routing method, as the initial routing method of DIAL since PeopleRank can be implemented without centralized infrastructures. Although there are many advanced routing methods beyond PeopleRank, PeopleRank is our ideal option since most advanced routing methods adopt centralized information in order to improve the routing performance. The basic routing process of PeopleRank is as follows.

1) Firstly, we consider vehicles are socially related to each other. Such social relationships can be based on explicit friendships on personal communication. Then, we adopt PageRank algorithm for calculating the centrality of different vehicles, which is called PeopleRank value in [2].
2) The PeopleRank value is given by

$$PeR(N_i) = (1-d) + d \sum_{N_j \in F(N_i)} \frac{PeR(N_j)}{|F(N_j)|} \quad (1)$$

where N_1, N_2,...,N_n are vehicles, $F(N_i)$ is the set of neighbors that link to N_i, and d is damping factor which is defined as the probability, at any encounter, that the social relation between the nodes helps to improve the rank of these nodes. This means that, the higher the value of d, the more the algorithm accounts for social relation between the vehicles. As a result, the damping factor is very useful in controlling the weight given to the social relations for the forwarding decision.
3) The PeopleRank value is updated every time when two vehicles encounter each other.
4) PeopleRank routing is a routing by continually selecting vehicles with higher PeopleRank values.

2) Building the address book: After we have chosen the initial routing method, in order to distributedly share the personalized information of vehicles, in the DIAL system, each vehicle maintains an address book by itself. For example, as shown in Figure 9, the address book stores Location Tables (LTables) and Contact Tables (CTables) of vehicles with ID 1, 2, 3, 4 and so on. A LTable records the frequently visited road ID and visited frequency of each road of the corresponding vehicle. A CTable records the frequently encountered vehicles and the encounter frequency of the vehicles of the corresponding vehicle. All the information can be used in improving the initial routing method introduced above in the information fusion based routing method. The calculation methods of visited frequency in LTable and encounter frequency in CTable are introduced as follows:

1) The encounter frequency of vehicles is measured by the method in [1]. Specifically, the contact utility is calculated every time when two vehicles encounter by:

$$C(v_i, v_j) = C_{old}(v_i, v_j) + (1 - C_{old}(v_i, v_j)) \times C_{init}(v_i, v_j) \quad (2)$$

where $C(v_i, v_j)$ is the updated encounter frequency utility; $C_{old}(v_i, v_j)$ is the old encounter frequency utility and $C_{init}(v_i, v_j)$ is the initial value of contact utility of all the vehicle pairs, which is set to a value selected from $(0, 1)$. This definition ensures that the two vehicles with a high encounter frequency have a larger encounter frequency utility.
2) The visited frequency of locations is measured by our previous method in [4]. The basic idea is as follows:
   a) First,, we divide road map to small road sections which can be denoted by road intersections.
   b) Then, each vehicle keeps recording the number of visiting times on each road sections.
   c) Finally, the top a few frequently visited locations can be identified.

3) Maintaining the address book: After vehicle A has calculated its own address information, vehicle B builds and maintains the address information of vehicle A as follows:

1) Once vehicle A delivers a packet to vehicle B, vehicle A delivers its address information to vehicle B with the packet.
2) Once vehicle B receives a packet and address information from vehicle A, vehicle B checks its address book. If vehicle A is in the address book, go to Step (3); otherwise, go to Step (4).
3) Vehicle B updates the address information of vehicle A by the new address information and changes the corresponding updated time to the current time stamp.
4) Vehicle B deletes the oldest address information from the address book and adds the address information of vehicle A, and sets the corresponding updated time to the current time stamp.

Besides building and maintaining its address book by the information provided by source vehicles, vehicle B also maintains its address book based on the information of other vehicles as follows:

1) Once vehicle B encounters another vehicle C, vehicle B checks whether there are useful address information of its frequently contact vehicles. If there is useful address information, go to Step (2).

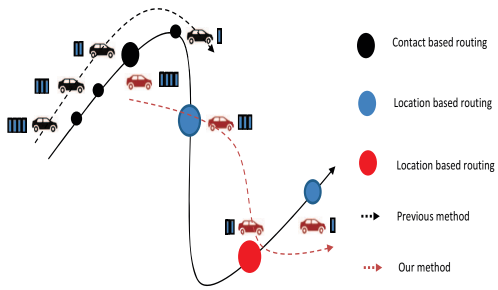


Fig. 10: Routing with combination of different methods

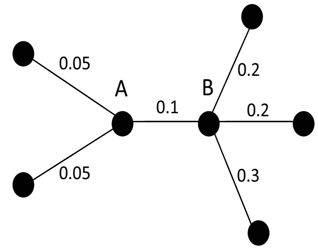


Fig. 11: An example of the routing process.

Vehicle ID	Thresholds		
	Contact	Centrality	Location
B	0.1	1.7	5
C	0.4	2.5	7
D	0.2	1.6	3
	......		

Fig. 12: An example of routing strategy table of vehicle A.

2) Vehicle B checks the corresponding updated time of the address information. If it is later than the address information stored in the address book, it updates the address information.

4) Selecting relay nodes based on address information: After vehicle A has built the address information of vehicle B in its address book, it chooses contact, centrality and location based routing methods simultaneously to deliver packets to B, which helps to find the best routing method. However, in a combination of different routing methods, if a relay node meets several vehicles all with higher contact utility first, then all the copies of the packet will be delivered by the same method only. Then, if the source vehicle meets a relay vehicle with very high centrality utility or location utility, the source vehicle will lose the chance to deliver the packet by those vehicles since each packet can only have limited copies in order to avoid the congestion. For example, as shown in Figure 10, a larger size of a circle presents a larger opportunity the source vehicle has to successfully deliver the packet. Suppose each packet can have 5 copies for routing and we continually send the copies to vehicles with any kind of utilities which are higher than current relay vehicle as shown as the trajectory of the black car in the figure. Then, all the copies are run out by relay vehicles with a little higher contact utilities (the relatively small black circles). Although the vehicle meet other vehicles with much higher location utility and centrality utility (the relatively big blue and red circles) later, it loses the chance to select them as relay vehicles. In order to overcome this problem, we set each method with a utility threshold. In addition to always forwarding packets to vehicles with higher utilities, the relay vehicles in each routing method should also have larger corresponding utility than its threshold.

To be more specific, we set three thresholds: centrality threshold, contact threshold and location threshold as follows:

1) **Contact threshold (Thr_{con}):** We define that the contact utility of the relay vehicle must be large than Thr_{con} for source vehicle to deliver a copy of a packet to the relay vehicle for contact based routing.
2) **Location threshold (Thr_{loc}):** We define that the frequently visited location between relay vehicle and target vehicle must be smaller than $\frac{1}{Thr_{loc}}$ for source vehicle to deliver a copy of a packet to the relay vehicle for location based routing.
3) **Centrality threshold (Thr_{cen}):** We define that the centrality utility of the relay vehicle must be large than Thr_{cen} for source vehicle to deliver a copy of a packet to the relay vehicle for centrality based routing.

As shown in Figure 10, in DIAL, we set thresholds to different methods which can guarantee that the copies can be left for those vehicles with very high utilities met later which are denoted as the bigger blue and red circles as shown as the trajectory of the brown car in the figure. Therefore, we can fuse the different kinds of information and improve the routing efficiency.

Once we set the thresholds for different routing methods, vehicle B can adopt different routing method to deliver packets to A simultaneously based on the following steps:

1) Once a packet is generated by vehicle B, vehicle B checks its address book. If the target vehicle A is in the address book, go to Step (3); otherwise, go to Step (2).
2) The current relay vehicle adopts the initial routing method to select next relay vehicle.
3) The current relay vehicle adds the location and contact information of vehicle A to the packet and then selects a vehicle as relay vehicle which has larger centrality utility than centrality threshold, larger contact utility than contact threshold or larger location utility than location threshold.
4) If a relay vehicle has more than one utility which is higher than the threshold, we pick the method which is used with the smallest number of times in order to balance the load caused by different methods.

B. Adaptive-learning Framework

In adaptive-learning framework, we first introduce a data structure named *routing strategy table* stored on each vehicle. The routing strategy table stores the thresholds of different utilities which are corresponding to different routing methods to different target vehicles. Then, we introduce the detailed method for maintaining the thresholds in the routing strategy table. Finally, we describe the whole DIAL routing process.

1) Building and maintaining routing strategy table: In the previous section, we set the centrality threshold, contact threshold and location threshold to constant values. Obviously, such a strategy is not ideal since optimal thresholds can be different from one vehicle pair to another vehicle pair. For

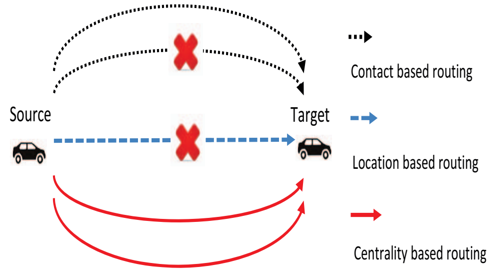


Fig. 13: An example of setting thresholds.

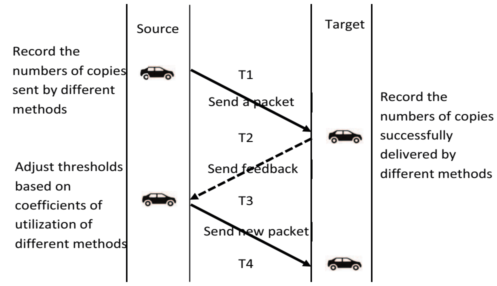


Fig. 14: An example of building and maintaining the routing strategy table.

example, as shown in Figure 11, there are two target vehicles A and B. Vehicle A is inactive and 0.1 is a high enough contact threshold for delivering packets to vehicle A, while vehicle B is active and 0.2 can be a good enough contact threshold for delivering packets to vehicle B. In order to set different thresholds for different vehicle pairs, in the adaptive-learning framework, a vehicle A maintains a routing strategy table as shown in Figure 12. The routing strategy table stores the thresholds for vehicle A to deliver packets to vehicles with IDs B, C, D and so on.

However, as we mentioned in Section III, it is difficult to predict the routing performances of different routing methods on different vehicle pairs. To solve this problem, in the adaptive-learning framework, instead of applying features of vehicle pairs to predict the performances of different routing methods on different vehicle pairs, we consider the routing process and all the features of vehicle pairs as a black box. By taking advantage of the human beings' communication feature that most interactions are generated by pairs of people who interacted often previously, the adaptive-learning framework tests the routing performances of different routing methods on different vehicle pairs. Then, based on the feedback of the tests, the adaptive-learning framework can adjust the thresholds of utilities for different routing methods.

To be more specific, the source vehicle A remembers the number of copies sent by different routing methods to target vehicle B. At the same time, target vehicle B remembers the number of copies successfully delivered by different routing method from source vehicle A and sends it back to source vehicle A with other packets. Then, vehicle A can calculate the success rates of different routing methods on itself. If the success rate of a routing method is higher than others, it means that the threshold of that utility is too low and a lot of copies delivered to vehicles with higher utility than the threshold are wasted. Therefore, we increase the threshold. Otherwise, if the success rate of a routing method is lower than others, it means that the threshold of that utility is too low. Therefore, we decrease the threshold. Here, we define the success rate of a specific method as:

$$SR_{con/loc/cen} = \frac{|SS_{con/loc/cen}|}{|S_{con/loc/cen}|} \quad (3)$$

where $SR_{con/loc/cen}$ is the success rate of contact, location and centrality based routing methods, respectively; $|S_{con/loc/cen}|$ is the set of copies sent by a specific method and $|SS_{con/loc/cen}|$ is the set of copies successfully delivered by a specific method. For example, as shown in Figure 13, the source node tries to send 5 copies of a packet, in which 2 of them are delivered by the contact based routing, 1 of them are delivered by the location based routing and 2 of them are delivered by the centrality based routing, respectively. Finally, 1 of them is successfully delivered by the contact based routing and 2 of them are successfully delivered by the centrality based routing, respectively. Therefore, we have $SR_{con} = \frac{1}{2} = 0.5$, $SR_{loc} = \frac{0}{1} = 0$ and $SR_{cen} = \frac{2}{2} = 1$.

Based on the calculated success rates, we adjust the thresholds of the three different methods as follows:

$$Thr^{new}_{con} = Thr^{old}_{con} + (SR_{con} - M) \times \Delta_{con} \quad (4)$$
$$Thr^{new}_{loc} = Thr^{old}_{loc} + (SR_{loc} - M) \times \Delta_{loc} \quad (5)$$
$$Thr^{new}_{cen} = Thr^{old}_{cen} + (SR_{cen} - M) \times \Delta_{cen} \quad (6)$$

where Thr^{new}_{con}, Thr^{new}_{loc} and Thr^{new}_{cen} are the new thresholds of contact, location and centrality based routing methods, respectively. Thr^{old}_{con}, Thr^{old}_{loc} and Thr^{old}_{cen} are the old thresholds of contact, location and centrality based routing methods, respectively. Δ_{con}, Δ_{loc} and Δ_{cen} are pre-defined constants for the new increments of thresholds of contact, location and centrality based methods, respectively, which we set them to 0.1. M is the median of the three success rates.

To sum up, as shown in Figure 14, the routing strategy table is built and maintained by the following steps:

1) Initially, vehicle A gives the thresholds to all the vehicles as constants.
2) When vehicle A sends a packet to vehicle B, vehicle A records the number of copies sent out by different methods, respectively, as shown in time $T1$ in Figure 14.
3) When vehicle B receives the copies of the packet sent by A, vehicle B records the numbers of copies successfully delivered to itself by different methods, respectively, as shown in time $T2$ in Figure 14.
4) When vehicle A receives the feedback sent by vehicle B, vehicle A adjusts the thresholds of different methods by Formula (4), Formula (5) and Formula (6), respectively, as shown in time $T3$ of Figure 14.

2) Detailed DIAL routing process: Based on the above description, we give a detailed DIAL routing process as follows:

1) Once a packet is generated by vehicle B, vehicle B checks its address book. If the target vehicle A is in the address book, go to Step (3); otherwise, go to Step (2).
2) The current relay vehicle adopts the initial routing method to select next relay vehicle. Then go to Step (5).
3) The source vehicle B checks its routing strategy table. If there is a routing strategy for target vehicle A, vehicle B uses the thresholds stored in the routing strategy table; otherwise, it uses the initial values of the thresholds. Then go to Step (4)
4) The current relay vehicle adds the location and contact information of vehicle A to the packet. Vehicles with larger centrality utility, larger contact utility and larger location utility than the corresponding (centrality, contact and location) thresholds are selected as relay vehicles.
5) The current relay vehicle checks its address book. If the target vehicle A is in the address book, go to Step (3); otherwise, go to Step (2).

V. Performance Evaluation

In this section, we evaluate the performance of DIAL and compare it with other methods. We conduct the trace-driven experiments based on both the *Roma* and *SanF* traces. In order to evaluate on continual interactions between vehicles in a long term, we recursively set the states of vehicles to the beginning of the trace data and replay the trace data once the trace data run out since the durations of the trace data are not long enough. Based on the above experiment environment, we use the following metrics to evaluate the routing performance:

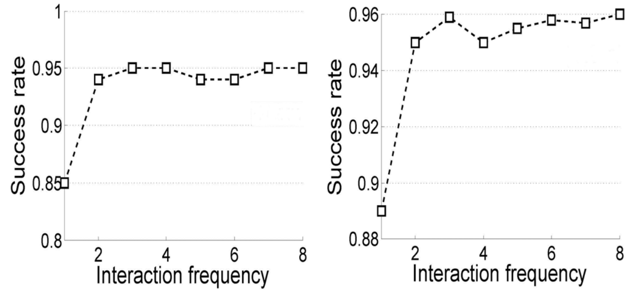
1) *Success rate:* The percentage of packets that successfully arrive at their target vehicles.
2) *Average delay:* The average time per packet for successfully delivered packets to reach their target vehicles.

Our evaluation is divided to two aspects:

1) From a micro-scope, we measure the routing performance of DIAL with different interaction frequencies since DIAL is designed based on the fact that most interactions are generated by pairs of people who interacted often previously.
2) From a macro-scope, we compare the performance of DIAL with the AAR [4], PeopleRank [2] and PROPHET [1] methods. AAR represents location based routing methods. PeopleRank represents centrality based routing methods. PROPHET represents contact based routing methods. The details of the methods are introduced in Section II.

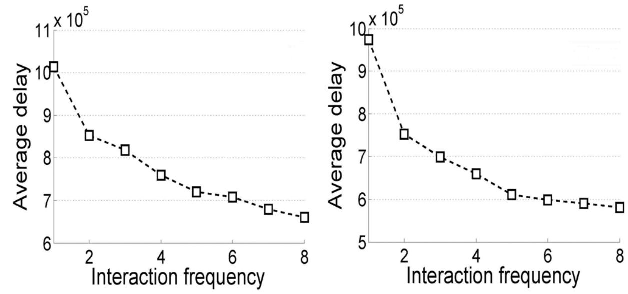
A. Performance Comparison With Different Parameters

It is obvious that DIAL is influenced by the interaction frequency. Therefore, we first analyze the influence of interact



(a) *Roma* (b) *SanF*

Fig. 15: The success rate vs. the interaction frequency.

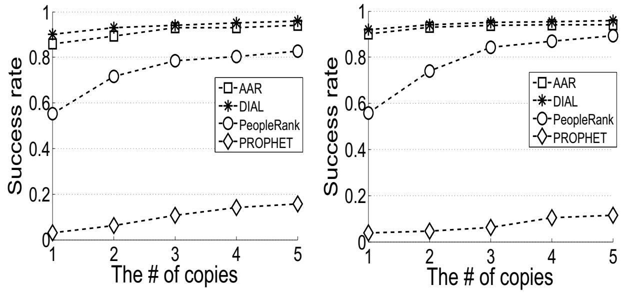


(a) *Roma* (b) *SanF*

Fig. 16: The average delay vs. the interaction frequency.

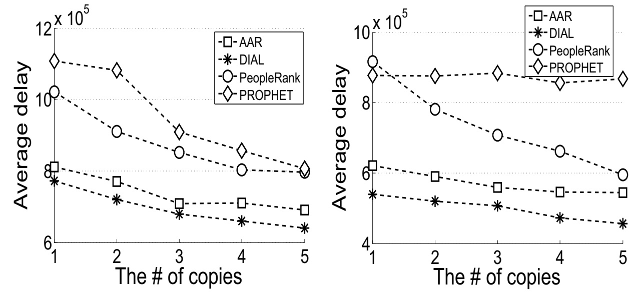
frequency on the performance of DIAL. In order to test the change of the routing performance when vehicle pairs continually interact with each other, we randomly select 100 vehicle pairs and continually generate packets between each of them. A vehicle pair A and B interact with each other one time means that A sends a packet to B and B sends a packet to A. Figure 15 shows the change of success rates and average delays of DIAL with the increasing of the times of the interactions between each vehicle pair. As shown in Figure 15, the success rate is significantly improved at the first several interactions. The reason is that, in the first time of interactions, the successfully delivered packets send the useful information from source vehicles to target vehicles. Then, the target vehicles send packets back to their source vehicles by taking advantage of the information came from the first interaction. Therefore, some of the vehicle pairs which initially cannot be delivered from target vehicle to source vehicle can successfully deliver the packet in the second time of interactions. Therefore, the success rate is significantly improved in the second time of interactions. By taking advantage of the information brought by first a few interactions, the routings on most vehicle pairs have been improved. Therefore, in the next following interactions, the success rate is not changing too much since routings on those vehicle pairs which the routings are failed in both sides cannot take advantage of DIAL.

On the contrary, the average delay continually decreases with the increase of the times of the interactions for a long time as shown in Figure 16. The reason is that, in the first a few times of interactions, by sending the location and contact



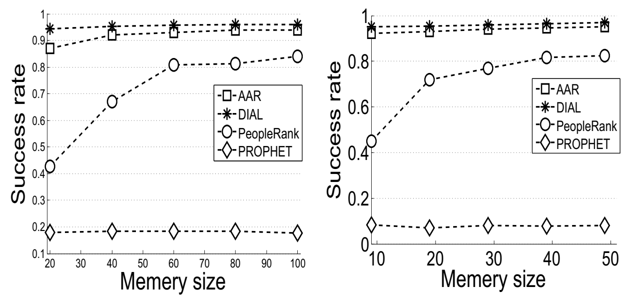
(a) *Roma* (b) *SanF*

Fig. 17: The success rate vs. different number of copies.



(a) *Roma* (b) *SanF*

Fig. 18: The average delay vs. different number of copies.



(a) *Roma* (b) *SanF*

Fig. 19: The success rate vs. different memory sizes.

information from one side to another side of the routings, the success rate and average delay can be improved obviously. In the subsequent interactions, although the routings on the vehicle pairs which the routings were failed in both sides cannot be improved, the vehicle pairs which are successfully interacted can still optimize their routing strategies based on our adaptive-learning framework. At the same time, for contact, location and centrality based routing methods, the routing method with higher success rate usually leads to a shorter average delay. Therefore, the average delay can be continually improved.

B. Performance Comparison With Other Methods

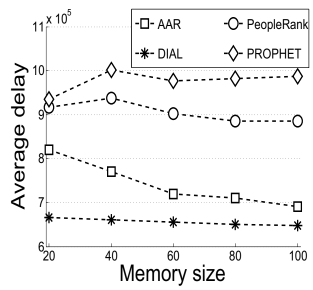
Then, we compare DIAL with other routing methods. In the evaluation of previous methods, the packets are randomly generated to evaluate and compare the performance of different methods. However, the interactions between people follows a special pattern. Therefore, in order to improve the accuracy of the evaluation, instead of randomly generating the packets among vehicles, we can reproduce a series of packets which follow the special pattern in [10], in which most interactions are generated by vehicle pairs that interacted often previously. Specifically, we define the probability to generate a packet between vehicles A and B as $\frac{r_{AB}}{\sum r}$, where r_{AB} is the rank of the vehicle pair of A and B and $\sum r$ is the sum of the ranks of all vehicle pairs. Initially, we set the rank of all vehicle pairs to 1. Then, every time when a packet is generated between A and B, the rank of the vehicle pair of A and B is increased by 1. As a result, most interactions are generated by vehicle pairs who interacted often previously.

1) Performance comparison with different numbers of copies: First, we compare the success rates and average delays with different numbers of copies of each packet. Figure 17 shows the success rates with different numbers of copies per packet based on the *Roma* and *SanF* traces, respectively. Generally, the performances follow DIAL>AAR>PeopleRank>PROPHET. The performance of AAR is better than PeopleRank since AAR uses the global information of each vehicle's frequently visited locations. DIAL performs slightly better than AAR since we take advantage of the human beings' communication feature and learn the best routing strategy for each vehicle pair. PROPHET performs the worst since it is difficult for a vehicle to encounter a vehicle that has a high probability to encounter the destination vehicles in the VDTNs. Although DIAL has similar performance as AAR, AAR uses the global information which is not easy to be implemented in reality. However, DIAL achieves similar success rate in a totally distributed way which is easy to be implemented in reality.

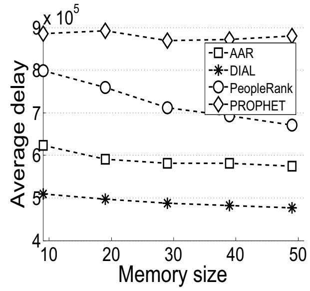
Figure 18 shows the average delays with different numbers of copies per packet based on the *Roma* and *SanF* traces, respectively. Generally, the average delays follow PROPHET>AAR>PeopleRank>DIAL. The delay of PROPHET is the largest since the relay vehicles need to wait for a long time to encounter a vehicle that has a high probability to encounter the destination vehicles in the VDTNs. The delay of DIAL is the smallest since we continually optimize the routing strategy for each vehicle pair during the routings.

Based on the above evaluation, we find that the success rate and average delay of DIAL are both improved. Although the improvement of success rate is not significant, DIAL does not rely on global information which makes it easier to be implemented in reality. Also, the average delay of DIAL is much shorter than the other methods.

2) Performance comparison with different memory sizes: Then, we compare the success rates and average delays with different memory sizes of the vehicles and we suppose 1 unit of memory can store 1 packet. Figure 19 and Figure 20 show the success rates and average delays with different memory sizes, respectively. Generally, the sensitivities of different methods to the memory sizes follow



(a) *Roma*



(b) *SanF*

Fig. 20: The average delay vs. different memory sizes.

PeopleRank>AAR>DIAL>PROPHET. The performance of PeopleRank is very sensitive to the memory size since all the packets tend to be forwarded to few vehicles with very high PeopleRank values and the limited memory size can significantly influence the routing process negatively. PROPHET is insensitive to the memory size, since the packets only tend to find those specific vehicles with high probability to encounter the target vehicles, which helps achieve load balance. However, PROPHET generates low success rate due to the same reason that the packets only tend to find those specific vehicles with high probability to encounter the target vehicle and therefore, it is difficult to encounter a suitable relay vehicle. AAR delivers a packet by many relay vehicles from one road intersection to another road intersection and if vehicles' memory is limited, some packets may lose the chance for delivery. Therefore, AAR is still relatively sensitive to the memory size. DIAL is relatively insensitive comparing with AAR and PeopleRank since we adopt the adaptive-learning framework which adjusts the routing strategies from time to time.

VI. Conclusion

In this paper, by fully taking advantage of the human beings' communication feature that most interactions are generated by pairs of people who interacted often previously, we proposed DIAL, an efficient routing method in VDTNs. DIAL has two components: information fusion based routing method and adaptive-learning framework. The information fusion based routing method enables DIAL to improve the routing performance by sharing and fusing multiple kinds of routing information without centralized infrastructures. Furthermore, based on the information shared by information fusion based routing method, the adaptive-learning framework enables DIAL to design a personalized routing strategies for different vehicle pairs without centralized infrastructures. Therefore, DIAL can not only share and fuse multiple kinds of routing information of each vehicle without centralized infrastructures, but also dynamically determine the personalized routing strategy for each vehicle pair. The trace-driven simulation demonstrates that DIAL can slightly improve the VDTNs routing success rate comparing with previous routing method which is based on centralized information. In our future work, we will try to let source vehicles utilize encounters with other vehicles in the network to pre-relocate the target vehicles in order to further improve the routing efficiency.

Acknowledgements

This research was supported in part by U.S. NSF grants NSF-1404981, IIS-1354123, CNS-1254006, IBM Faculty Award 5501145 and Microsoft Research Faculty Fellowship 8300751.

References

[1] A. Lindgren, A. Doria, and O. Scheln, "Probabilistic routing in intermittently connected networks.," *Mobile Computing and Communications Review*, vol. 7, pp. 19–20, 2003.

[2] A. Mtibaa, M. May, C. Diot, and M. H. Ammar, "Peoplerank: Social opportunistic forwarding.," in *Proc. of INFOCOM*, pp. 111–115, IEEE, 2010.

[3] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant manets.," in *Proc. of MobiHoc*, pp. 32–40, ACM, 2007.

[4] B. Wu, H. Shen, and K. Chen, "Exploiting active subareas for multi-copy routing in vdtns," in *Proc. of ICCCN*, pp. 1–10, IEEE, 2015.

[5] I. Leontiadis and C. Mascolo, "Geopps: Geographical opportunistic routing for vehicular networks.," in *Proc. of WOWMOM*, pp. 1–6, IEEE, 2007.

[6] J. B. Link, D. Schmitz, and K. Wehrle, "Geodtn: Geographic routing in disruption tolerant networks.," in *Proc. of GLOBECOM*, pp. 1–5, IEEE, 2011.

[7] K. Chen and H. Shen, "Dtn-flow: Inter-landmark data flow for high-throughput routing in dtns.," in *Proc. of IPDPS*, pp. 726–737, IEEE, 2013.

[8] J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine, "Maxprop: Routing for vehicle-based disruption-tolerant networks.," in *Proc. of INFOCOM*, IEEE, 2006.

[9] T. Hossmann, T. Spyropoulos, and F. Legendre, "Know Thy Neighbor: Towards Optimal Mapping of Contacts to Social Graphs for DTN Routing.," in *Proc. of INFOCOM*, pp. 866–874, IEEE, 2010.

[10] A. G. Miklas, K. K. Gollu, K. K. Chan, S. Saroiu, K. P. Gummadi, and E. de Lara, "Exploiting Social Interactions in Mobile Systems," Lecture Notes in Computer Science, pp. 409–428, Springer, 2007.

[11] B. Lorenzo, B. Marco, L. Pierpaolo, B. Giuseppe, A. Raul, and R. Antonello, "CRAWDAD data set roma/taxi (v. 2014-07-17)." Downloaded from http://crawdad.org/roma/taxi/, July 2014.

[12] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "CRAWDAD data set epfl/mobility (v. 2009-02-24)." Downloaded from http://crawdad.org/epfl/mobility/, Feb. 2009.