DEBUT: Delay Bounded Service Discovery in Urban Vehicular Ad-hoc Networks

Fenggang Wu, Hongzi Zhu, Jia-Liang Lu, Min-You Wu
Department of Computer Science and Engineering Shanghai Jiao Tong University
Email: fenggang.wu@sjtu.edu.cn, {hongzi,lu-jl,wu-my}@cs.sjtu.edu.cn

***Abstract*—This paper studies delay-bounded service discovery in urban Vehicular Ad-hoc Networks (VANETs), which refers to locating resources and services (e.g., local sensor data and multimedia content) distributed on individual vehicles in the network within a certain delay bound. To facilitate the discovery process, a set of vehicles, called service directories (SDs), can be selected to store the index information of all the resources in the network. Selecting an optimal SD set with minimal size while satisfying the users' requirement of a bounded query response delay is very difficult due to the disruptive nature of VANETs. In this paper, we formulate the Delay Bounded Service Directory Selection (DB-Sel) problem as an optimization problem that minimizes the number of SDs under the delay bound constraint. We prove theoretically that the DB-Sel problem is NP-Complete even when the future positions of vehicles are known a priori. We observe and prove that the number of vehicles encountered by arbitrarily selected SDs within a given delay follows a normal distribution. We also find the contact probabilities among the vehicles exhibit strong temporal correlation. With these observations, we develop a heuristic algorithm which iteratively selects the best candidate according to the normal distribution property and the historical contact probability. We prove that our algorithms have a guaranteed performance approximation ratio compared to the optimal solution. Extensive trace-driven simulation results demonstrate that our algorithm can guarantee the required query delay and select SD sets 20% smaller than those selected by alternative algorithms.**

I. INTRODUCTION

Urban *Vehicular Ad-hoc Networks* (*VANETs*) are networks comprised of many urban vehicles which are equipped with wireless communication modules and can communicate with each other without the aid of any infrastructure. This emerging technology has enabled many potential applications, such as emergency handling applications, traffic information service, and other safety or infotainment applications. Among those applications, service provisioning among vehicles is an appealing application where individual vehicles can share local data (e.g., sensor data and multimedia content) or resources (e.g., storage, access to the Internet) to other vehicles in the network.

In order to realize this application, such local resources must be discovered first before being used. We refer to locating distributed resources and services in the network within a certain delay bound as *delay-bounded service discovery*. To provide this in VANETs, however, is very challenging for the following three reasons. First, distributed service discovery schemes are more appropriate. One possible solution is to use a server to index and maintain the information about resources on individual vehicles. Queries can be sent to the server to locate the requested resources in the network. Such centralized schemes are simple but not feasible in practice due to the single point of failure and the prohibitive constant communication cost (e.g., via 2G/3G networks) for updating and querying the information on the distributed resources. A better solution might store the index information on vehicles, leveraging free vehicle-to-vehicle communications. Second, the cost of updating the index information of resources should be minimized. Since vehicles use wireless communications and the bandwidth of wireless channels is quite limited, it is of great importance to reduce the communication cost for updating. Besides, the deployment cost to equip the indexing vehicles must be reduced. Last, queries for certain resources are generally time-constrained. Long query response delay may cause an inconsistent view of the required resources. For example, a user may query for a certain amount of storage in the network. If the response to the query takes a long period of time, the current amount of available storage may have changed during this period and cannot satisfy the query anymore. In VANETs, the disruptive nature of the network may incur a significant delay, which makes fast service discovery very hard.

In the literature, the *Decentralized Directory-based Architecture* [1][2] has been proposed to solve the service discovery problem in mobile ad-hoc networks (MANETs), in which some nodes in the network are selected as *Service Directories* (*SDs*), to store the service description for those nodes with shared resources (called *service providers*). Nodes which are requesting certain services in the network (called *service requesters*) send queries to SDs. In response, SDs return the description of matching services. However, these existing schemes are originally designed for MANETs and do not take the significant transmission delay into consideration. As a result, there is no existing solution, to the best of our knowledge, to tackle the delay-bounded service discovery problem in VANETs.

In this paper, we propose an innovative service discovery scheme, called *DEBUT*, which, with a given probability, can guarantee that queries issued from a required proportion of vehicles can be answered within a given delay bound while minimizing the number of vehicles which serve as SDs, so as to minimize the bandwidth for updating service descriptions

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

on the SDs and the deployment cost for equipping such SD vehicles with more powerful processing and communicating modules. Specifically, we define the *delay bounded coverage ratio* , as the fraction of vehicles that can be served by SDs within the required delay bound and and the *delay bounded guarantee ratio* as the confidence interval to guarantee this coverage ratio. Then we formulate the *delay bounded service directory selection* (*DB-Sel*) problem as an optimization problem that minimizes the number of SDs while ensuring the delay bounded coverage ratio and the delay bounded guarantee ratio requirements mentioned above. We prove this optimization problem is NP-Complete even when the future positions of vehicles are known a priori. We further propose a heuristic iterative algorithm when all future positions of vehicles are known.

For more realistic settings where the future positions of the vehicles cannot be known in advance, we make an empirical study of real Global Positioning System (GPS) trace data of over 4,000 operational taxis in Shanghai. We observe and prove that the number of vehicles which can encounter any vehicle in a given SD set follows a normal distribution. In addition, we also observe that the contact probabilities among vehicles exhibit strong temporal correlation. Based on these observations, we design a greedy algorithm to iteratively add the vehicle that can maximize the expected increment of the fraction of vehicles that can encounter the SD vehicles within the delay bound. We give a provable performance approximation ratio guarantee for both algorithms compared to the optimal solution. Extensive trace-driven simulations demonstrate the efficiency of the DEBUT design. The results show that our algorithm can guarantee the required query delay and select SD sets 20% smaller than those selected by alternative algorithms.

Our main contributions can be highlighted as follows:

- We formally define the DB-Sel problem to address the fundamental SD selection issue in the SDP design in VANETs. Also we prove DB-Sel problem is NP-Complete even when the future vehicular positions are assumed to be known a priori.
- In empirical study of real traces we observe that the number of vehicles encountered by an SD set follows a normal distribution, and that the contact probabilities among individual vehicles exhibit strong temporal correlation. Also by analytical deduction we confirm the normal distribution observation. Based on these observations and analysis we propose greedy algorithms to solve the DB-Sel problem with provable performance approximation ratio compared to optimal solutions.
- We conduct extensive simulation based on the real GPS traces of over $4,000$ operational taxis and find that our algorithms outperform alternative algorithms by selecting an SD set 20% smaller while meeting the delay bound constraint.

The remainder of this paper is organized as follows. Section II discusses related work. In Section III we present the system model and formally define the problem under consideration. The analysis of the *delay bounded cover set* is presented in Section IV. Algorithm design is presented in Section V. In Section VI we present the evaluation methodology as well as the comparison results. We conclude the paper in Section VII. A preliminary work was published in [3], and this is the complete version.

II. RELATED WORK

According to whether a *directory* exists or not, the service discovery architecture can be divided into two categories: *directory-based* approaches and *directory-less* approaches. In the directory-less approaches, the service providers do not spread the service description to other nodes [4][5]. In [4][5] service servers periodically push service advertisements to their 1-hop neighbors by broadcasting. In the centralized directory-based approaches [6][7], one or a few service directory node(s) are selected to store the service descriptions of all the available service in the whole network. In the decentralized directory-based approaches [2][8], the directories are distributed to cope with the availability and scalability problem. In [2], directories are deployed so as to ensure at least one directory is reachable within a fixed number of hops. In summary, our work differs from the existing work in that we take the significant delay of VANET into consideration when designing the SDP for the network.

III. SYSTEM MODEL

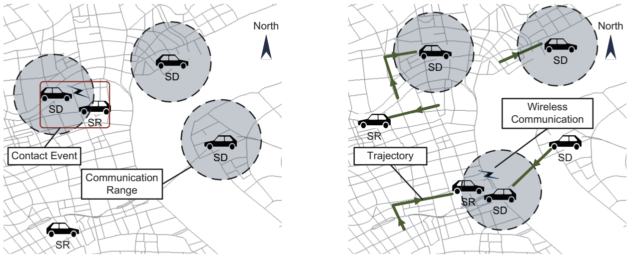
In this section, we first describe the preliminaries and assumptions. Then we give a formal definition of the *Delay Bounded Service Directory Selection (DB-Sel)* problem.

A. Preliminaries and Assumptions

In VANETs, two vehicles can communicate with each other when they come within the *communication range* R and we refer to such a communication opportunity as a *contact event*. The n vehicles in the VANETs are denoted by $\mathcal{U}$. The set of all the location sequences of the vehicles in $\mathcal{U}$ between time t_a and t_b is denoted by $L(t_a, t_b)$.

Every vehicle v_i is assumed to have some kinds of services Γ_S. Each of the services Γ_S has a service description file $\mathcal{F}_S$ that describes the details of Γ_S, including the service name, type, the service provider's ID, and expiry time. Each v_i may also want to discover some kinds of services by sending a service query message Ψ_Q that includes the key words of the service it is interested in and the requester's ID.

A service provider vehicle has to register its new or updated service Γ_S on the nearest SD by sending the corresponding service description file $\mathcal{F}_S$. Then this SD will be in charge of spreading $\mathcal{F}_S$ to all the other SDs. When a service requester vehicle wants to access a certain type of service, it will send a service query message Ψ_Q to the first SD it encounters, and the SD will return the matching service description file $\mathcal{F}_S$ to the requester. Then the service requester decides which service to invoke and contacts the service provider directly using the existing routing protocols (Fig.1).



(a) Service requester (SR) wants to query SDs who store the service descriptions for service providers.

(b) After some delay, the query of service requesters (SRs) can be answered when meeting SDs.

Fig. 1. An illustration of delay bounded service discovery in VANETs.

We assume the query will fail if it has not been answered within a time bound D. At the same time, the number of vehicles in the set S cannot be unlimited because 1) the cost for storing and updating the index information of services should be minimized, since the bandwidth of a wireless channel is quite limited, and more SDs incur more communication cost for updating service descriptions, and 2) such directory vehicles will have a cost for deployment (for example, directory vehicles may need better communication modules, larger memories and more powerful processors). For a similar reason, we assume the service requesters only contact the SDs by one-hop communication to reduce the service discovery communication overhead and save the precious bandwidth in VANETs.

B. Problem Formulation

In this section, we define the DB-Sel problem as an optimization problem that minimizes the size of the SD set under the delay bound constraint. And then we show that the DB-Sel problem is NP-Complete even when future traces are known a priori.

Vehicle v is said to be covered by set S with a delay D if there is any $u \in S$ that can encounter v within delay D. It is denoted as $v \in \mathcal{C}(S,D)$, where $mathcalC(S,D)$ is called the *delay bounded cover set*.

Definition 1 (Delay Bounded Coverage Ratio)**.** *The delay bounded coverage ratio is the proportion of the vehicles $v \in \mathcal{U}$ that satisfy $v \in \mathcal{C}(S,D)$. It is denoted as*

$$\eta(S,D) \triangleq \frac{\| \mathcal{C}(S,D) \|}{\| \mathcal{U} \|}, \tag{1}$$

where $\| \centerdot \|$ means the size of a set.

Without loss of clarity, we call it the delay bounded coverage ratio *coverage ratio* for short. Here we arrive at the formal definition of the DB-Sel problem.

Definition 2 (Delay Bounded Service Directory Selection Problem)**.**

$$\begin{aligned} \min_{S} \quad & \| S \| \\ s.t. \quad & \eta(S,D) > \gamma \quad , \end{aligned} \tag{2}$$

where γ is a required bounded coverage ratio.

Further, according to whether the future trace is given, we can have two versions of the DB-Sel problem, i.e., the *Determined DB-Sel* (DDB-Sel) problem and *Non-determined DB-Sel* (NDB-Sel) problem.

Definition 3 (Determined Delay Bounded Service Directory Selection Problem)**.** *The determined delay bounded service directory selection problem is the DB-Sel problem with the future vehicle trajectories $L(t, t+D)$ given a priori.*

Theorem 1. *The DDB-Sel problem is NP-Complete.*

The proof is omitted due to the page limit and can be found in our technical report [9].

In realistic settings, however, we can only make decisions based on historical trajectory data $L(t-T, t)$, and the future positions of the vehicles are difficult or even impossible to determine. Therefore given a set of directories S, the number of covered vehicles $\| \mathcal{C}(S,D) \|$ will fluctuate over time and is no longer a determined value. We denote the size of the covered vehicles as a random variable $\xi(S,D)$. Subsequently, the coverage ratio $\eta = \frac{\xi(S,D)}{\|\mathcal{U}\|}$ also becomes a random variable. We cannot guarantee η to be above the required coverage ratio γ at all time. Thus we change the constraint into a stochastic one by introducing the *delay bounded guarantee ratio*.

Definition 4 (Delay Bounded Guarantee Ratio)**.** *The delay bounded guarantee ratio is the probability that the delay bounded coverage requirement $\eta(S,D) > \gamma$ is satisfied.*

$$\varrho(S,D,\gamma) \triangleq \Pr\{\eta(S,D) > \gamma\}. \tag{3}$$

Without loss of clarity, we call it the delay bounded guarantee ratio *guarantee ratio* for short. Here, we give the formal definition of the non-determined DB-Sel problem as follows:

Definition 5 (Non-determined Delay Bounded Service Directory Selection Problem)**.**

$$\begin{aligned} \min_{S} \quad & \| S \| \\ s.t. \quad & \varrho(S,D,\gamma) > \varpi \end{aligned} \tag{4}$$

where ϖ is a required guarantee ratio*, given a period of historical vehicle trajectories $L(t-T, t)$.*

IV. DELAY BOUNDED COVER SET ANALYSIS

In this section, in order to design an algorithm that satisfies the constraint in eq.4, we study the statistical characteristics of the variable $\xi(S,D)$. First we prove that $\xi(S,D)$ follows a normal distribution, then we verify this by empirical study. This empirical study also shows that the contact probabilities among vehicles exhibit strong temporal correlation.

A. Analytical Study for $\xi(S,D)$

In this section, we give the following theorem with the proof [9] omitted due to the page limit.

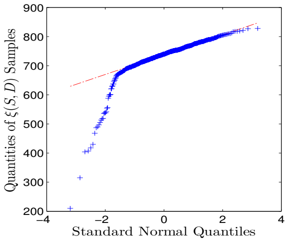


Fig. 2. Q-Q plot of $\xi(S, D)$ samples vs. standard normal. $|S| = 200$, $D = 30$min.

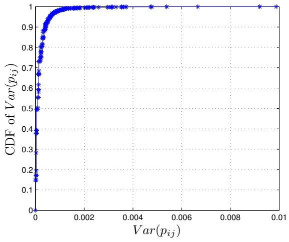


Fig. 3. The CDF of the variance of all the pairwise contact probabilities p_{ij}.

Theorem 2. *Random variable $\xi(S, D)$ follows a normal distribution when the number of vehicles goes to infinity.*

$$\xi(S,D) \sim \mathcal{N}\left(\sum_{i=1}^{n} q_i(S,D), \sum_{i=1}^{n} q_i(S,D)(1 - q_i(S,D))\right), \tag{5}$$

where $q_i(S, D) \triangleq \Pr\{v_i \in \mathcal{C}(S, D)\}$.

B. Empirical Study for $\xi(S, D)$

We perform extensive statistical analysis and find that $\xi(S, D)$ follows a normal distribution at a wide range of parameters, which verifies the conclusion of Section IV-A. Also we find that the contact probabilities have strong temporal correlation over time.

Our empirical study is based on a large dataset of real GPS traces of Shanghai taxis collected from more than $4,000$ operational taxis over a two-year period from Jan. 2006 to Dec. 2007. Every report in the trace files includes the vehicle ID, the GPS position, the direction of movement, the speed and a time stamp. By selecting a directory set S and a delay bound D, $\xi(S, D)$ can be calculated based on the traces.

For an arbitrarily selected SD set S, and a given delay bound D, we collect $\xi(S, D)$ samples and make a Q-Q plot versus the standard normal distribution. The results show that $\xi(S, D)$ exhibits significant characteristics of a normal distribution. One example is given by Fig.2.

The second observation is that the contact probability exhibits a strong temporal correlation. We choose two consecutive time periods (two weeks each) and calculate the pairwise contact probability p_{ij} between any two vehicles v_i and v_j. Here all the p_{ij} can be viewed as samples of a random variable and the entropy of p_{ij} can be computed from its probability distribution. We find that the entropy of the p_{ij} in the second time period is reduced by about 80% if given the probability distribution of the first time period, which implies that the historical data can greatly help to reduce the uncertainty of the future contact probability. Furthermore we find the variance of p_{ij} over time is very small. We calculate the variance of all the p_{ij} along the whole month and plot the CDF of the variance (Fig.3). We can see that over 95% of the vehicle pairs have a contact probability variation less than 0.001, meaning that the majority of the pairs have a stable contact probability which implies a strong temporal correlation of the contact probabilities, so that the future contact probabilities can be predicted well from historical contact probabilities.

Another finding is that the probability that a vehicle v_i is covered by all the vehicles u in set A, denoted by $p_i(A, D)$, drops below 0.1% when $\parallel A \parallel \geq 3$ and can be neglected compared to the sets A with smaller sizes.

Since $\xi(S, D)$ is normally distributed, given the parameters of the distribution, μ and σ^2, we can determine whether the requirement defined by eq.4 is satisfied based on the CDF of a standard normal distribution. Meanwhile, because the contact probabilities have strong temporal correlation, we can make good predictions of the probability of future contacts. This gives us a hint for designing the history-based iterative algorithm in the next section.

V. DELAY BOUNDED SDP ALGORITHMS

In this section, due to the NP-Completeness of DDB-Sel, we first give a greedy algorithm for the DDB-Sel problem, then develop it into a new heuristic iterative algorithm that deals with the NDB-Sel problem according to the observations in Sec.IV. Finally we prove that our algorithms have a guaranteed performance approximation ratio compared to the optimal solution.

A. Algorithm for DDB-Sel Problem

Due to the NP completeness of DDB-Sel, it is difficult to find the optimal solution, but we can design heuristic algorithm to solve this problem.

Since the objective of the algorithms is to achieve a required coverage ratio γ using as few SDs as possible, we can follow the heuristic of adding the vehicle which has the most uncovered neighbors. This leads to a simple greedy iterative algorithm.

According to the assumption that the future positions of the vehicles are known, the algorithm first calculates all the contact events during the future D time according to $L(t, t + D)$ and constructs a graph consisting of all the vehicles as vertices and all the edges indicating that a pair of vehicles have contacts in the next time interval of D. Then the algorithm iteratively adds the vehicle that can cover the most uncovered neighbors into the SD set, until the total number of covered vehicles satisfies the γ requirement. The pseudocode of this algorithm is omitted due to the page limit. It is available in our technical report [9].

B. Algorithm for NDB-Sel Problem

Based on the analysis in Sec.IV, we make improvements to the greedy iterative approach in the DDB-Sel problem and develop it into a heuristic algorithm to deal with the NDB-Sel problem. Before making this algorithm migration, we have to answer two key questions:

- how to select the best SD candidate in each iteration, and
- when to stop the iteration process.

For the first question, our algorithm will select the next candidate with the maximum expected *coverage ratio gain*.

As is shown in eq.5, the expected size of the delay bounded cover set μ can be estimated by $\frac{1}{n}\sum_{i=1}^{n} q_i$. When a new vehicle is added to S, all q_i will increase correspondingly. The expected coverage ratio gain g_i on v_i when adding new vehicle v_j into the S is $g_i(S, D, v_j) = q_i(S \cup v_j, D) - q_i(S, D)$, where $q_i(S, D) = \sum_{i=1}^{\infty}(-1)^{(i+1)} \sum_{A \in S, \| A \| = i} p_i(A, D)$.

It is not feasible in practice to use this equation for calculating $q_i(S, D)$. However, since $p_i(A, D)$ can be neglected when $\| A \| \geq 3$ (Sec.IV-B), $q_i(S, D)$ can be estimated by $\tilde{q}_i(S, D) = \sum_{v_k \in S} p_i(\{v_k\}, D) - \sum_{v_k, v_l \in S, k<l} p_i(\{v_k, v_l\}, D)$. Therefore the gain $g_i(S, D, v_j)$ can be estimated by $\tilde{g}_i(S, D, v_j) = \tilde{q}_i(S \cup v_j, D) - \tilde{q}_i(S, D) \ = p_i(\{v_j\}, D) - \sum_{v_k \in S} p_i(\{v_j, v_k\}, D)$, which means the gain of the coverage ratio on v_i when adding v_j in to S can be estimated by the probability of v_j covering v_i minus the sum of the co-cover probabilities of all the $\{v_j, v_k\}$ co-covers v_i, where v_k is an element in S. Then the gain of μ for adding v_j can be approximated by $\tilde{g}(v_j) = \sum_{v_i \notin S} \tilde{g}_i$, where $\tilde{g}(v_j)$ can be calculated based on historical data (Sec.IV-B).

For the second question, we will use the normal distribution property as well as a check based on the historical data to determine when to stop the iteration.

The algorithm estimates the normal distribution parameters μ and σ^2 of $\xi(S, D)$ by eq.5. Then, according to the CDF of the standard normal distribution, the algorithm can check whether the required coverage ratio and guarantee ratio are both satisfied. We refer to this as the *Gaussian Check*.

Besides the Gaussian Check, to mitigate the possible inaccuracy caused by estimating $\sum \tilde{g}_i$, the algorithm also performs a second check to make sure that eq.4 is satisfied, namely the *Historical Check*. In the Historical Check, the algorithm splits the historical data in past T time by a time interval of D. So there are $N_a = \frac{T}{D}$ time intervals in total. Given an SD set S, the algorithm counts the times when the coverage ratio requirement γ is satisfied, which is denoted by N_s. Then ϖ is estimated by $\varpi^* = \frac{N_s}{N_a}$. The Historical Check is satisfied when $\varpi^* > \varpi$. The algorithm stops and returns the directory set when both of them are passed. The pseudocode of this algorithm is omitted due to the page limit and is available in the technical report [9].

C. Approximation Ratio of the Algorithms

In this section, we show that our algorithms for the DDB-Sel and NDB-Sel problems have provable approximation ratios compared to the optimal solutions. We have the following theorem with the proof omitted due to the page limit. The proof is available in [9].

Theorem 3. *Let A be the solution selected by the iterative algorithm of DDB-Sel or NDB-Sel and B^* be the set with the optimal solution of the corresponding DB-Sel problem. Then $\| A \| \leq (1 + \frac{k_0}{(e-1)k_m}) \| B^* \|$, where k_0 and k_m is the incremental value of the expected covered vehicle number when adding the first and the last element in A.*

VI. PERFORMANCE EVALUATION

In this section, first we present the methodology and the experimental setup of the performance evaluation. Then we compare the performance of our proposed algorithms with alternative algorithms.

A. Methodology and Experimental Setup

We conduct a simulation based on real traces to show the efficiency of our proposed algorithms. The traces are described in Sec.IV-B.

The metric for the performance evaluation is the size of S. We take the controlling variables of γ, ϖ and the delay bound D as the influencing factors. The default settings are: $D = 60$ min, $\gamma = 0.95$, $\varpi = 0.9$. The communication range is set to 500m.

B. Compared Algorithms

It is computationally infeasible to get the optimal solution. Also, to the best of our knowledge, no existing approach can be directly adapted to our VANET scenario. So we choose some alternative algorithms as the baseline for comparison.

Random Selection (Random) This algorithm iteratively adds new randomly chosen vehicles to the SD set, and performs a Historical Check to determine when to stop the iteration.

Max Prob. Degree (MAX-DEG) It uses historical information to get the pairwise contact probabilities. For each vehicle v, it computes the sum of all the contact probabilities between v and the neighboring vehicles. It iteratively adds the vehicle with the maximum probability sum to the SD set, until the SD set satisfies the Historical Check.

C. Results

We investigate the impact of the required coverage ratio on the SD selection performance of the algorithms (Fig.4). We can see that the DDB-Sel algorithm produces a much smaller SD set than other algorithms as expected, which means knowing the future locations of the vehicles will help to choose much better SD candidates. On average the NDB-Sel algorithm selects an SD set about 20% smaller than Random and about 40% smaller than MaxDeg. As the required coverage ratio increases, the size of the SD set increases accordingly. This is reasonable in that more SDs are needed to achieve better coverage of the vehicles. We can see that the Random algorithm can even achieve a better performance than MaxDeg. One possible explanation is that some vehicles having close social relations or geographical vicinity may form into groups and they have much higher contact probabilities among themselves than between other vehicles pairs. Therefore MaxDeg tends to choose the vehicles in such groups who have large contact probability sum. However, such candidates have a small increment value of coverage ratio because the neighbors of the newly added vehicle have largely been covered by previous added vehicles from this group with high probability. Our NDB-Sel algorithm takes this issue into account and

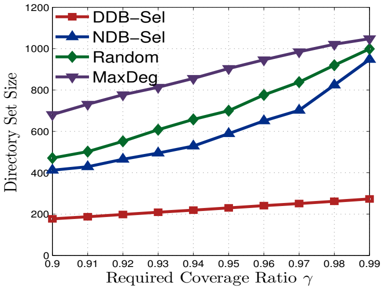


Fig. 4. SD set size vs. required coverage ratio. $\varpi = 0.9$, $D = 60$min.

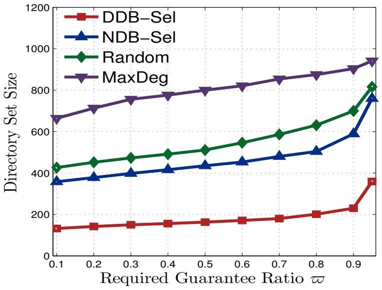


Fig. 5. SD set size vs. required guarantee ratio. $\gamma = 0.95$, $D = 60$min.

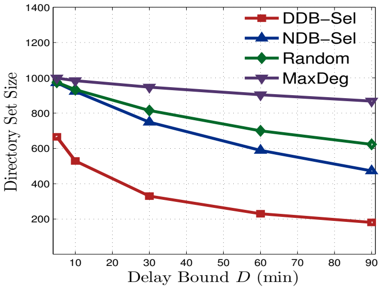


Fig. 6. SD set size vs. required delay bound D. $\gamma = 0.95$, $\varpi = 0.9$.

selects SDs according to the incremental expected coverage ratio and thus can have better performance.

Then we study the impact of the required guarantee ratio on the performance of the algorithms (Fig.5). We can see DDB-Sel still produces the SD sets with the smallest sizes. Even when the required coverage ratio and guarantee ratio are set to high probabilities as 0.95 and 0.9, a small SD set of about only 200 vehicles would be sufficient to meet such requirements. Again, NDB-Sel gives smaller SD sets than Random or MaxDeg. Besides we find that required coverage ratio has a more significant influence on the SD size than the required guarantee ratio. We compare Fig.4 and Fig.5, and see that the SD size increases more quickly when the required coverage ratio increases in the interval of $[0.9, 0.95]$ than when the required guarantee ratio increases in a much larger interval of $[0.1, 0.95]$. This is possibly because the number of covered vehicles over time has a small variation. Therefore increasing the number of SDs can change the actual guarantee ratio rapidly when it is close to the required guarantee ratio, but it cannot increase the actual coverage ratio so quickly to reach a higher required coverage ratio.

Finally we investigate the impact of the delay bound on the performance of the algorithms. From Fig.6 we see that again DDB-Sel algorithm selects SD sets about 50% smaller than other algorithms as expected. Also we find that NDB-Sel outperforms Random and MaxDeg by producing smaller SD sets. The directory set size decreases as the delay bound increases, which is reasonable because the contact probability increases when a larger delay bound is permitted, and therefore a smaller SD set will cover more vehicles. Another observation is that as the delay bound goes up, the advantage of NDB-Sel over Random and MaxDeg becomes increasingly significant. This is because the contact events exhibit greater statistical regularity with longer observation time duration.

VII. CONCLUSION AND FUTURE WORK

In this paper, we have studied the Delay Bounded Service Discovery Selection (DB-Sel) problem in Urban VANETs. The objective is to minimize the directory size while satisfying the coverage ratio and guarantee ratio requirement. First we proved theoretically that this problem is NP-Complete even under the assumption of prior knowledge of the future vehicles traces. Then to deal with the more realistic situation where future traces are unknown, we observe and prove that the size of the delay bounded cover set follows a normal distribution, and also observe that the contact probabilities exhibit strong temporal correlation over time. Then based on this analysis and observations we propose a greedy iterative algorithm to tackle the DDB-Sel problem and improve it into an algorithm that addresses the NDB-Sel problem. We prove that our heuristic algorithms have a guaranteed performance approximation ratio compared to the optimal solutions. Extensive trace-driven simulation shows our algorithms select SD sets 20% smaller than those of the alternative algorithms. Our future work may include: considering the multi-hop communication between vehicles in service discovery and selecting better SD sets by predicting future vehicular trajectories.

ACKNOWLEDGMENT

This research was supported by NSF of China under grant No.60773091, No.61100210 and Science and Technology Committee of Shanghai Municipal under grant No.12dz1507400.

REFERENCES

[1] U. C. Kozat and L. Tassiulas, “Service discovery in mobile ad hoc networks: an overall perspective on architectural choices and network layer support issues,” *Ad Hoc Networks*, vol. 2, pp. 23–44, 2004.

[2] F. Sailhan and V. Issarny, “Scalable service discovery for manet,” in *Pervasive Computing and Communications, 2005. PerCom 2005. Third IEEE International Conference on*. IEEE, 2005, pp. 235–244.

[3] F. Wu, H. Zhu, J. Lu, and M.-Y. Wu, “On optimal service directory selection in urban vehicular ad-hoc networks,” in *ACM CoNEXT Workshop on Urban Networking, UrbaNe*, 2012.

[4] M. Nidd, “Service discovery in DEAPspace,” *IEEE Wireless Communications*, vol. 8, pp. 39–45, 2001.

[5] R. Hermann, D. Husemann, M. Moser, M. Nidd, C. Rohner, and A. Schade, “DEAPspace - Transient ad hoc networking of pervasive devices,” *Computer Networks*, vol. 35, pp. 411–428, 2001.

[6] J. Waldo, *The Jini Specifications*, 2nd ed., K. Arnold, Ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2000.

[7] E. Guttman, C. Perkins, J. Veizades, and M. Day, “Service location protocol, version 2,” 1999.

[8] V. Raychoudhury, J. Cao, W. Wu, Y. Lai, C. Chen, and J. Ma, “K-directory community: Reliable service discovery in MANET,” *Proc. IEEE PERCOM*, vol. 7, pp. 140–158, 2011.

[9] F. Wu, H. Zhu, J. Lu, and M.-Y. Wu, “Technical report of debut: Delay bounded service discovery in urban vehicular ad-hoc networks,” Shanghai Jiao Tong University, Tech. Rep., 2012. [Online]. Available: http://wirelesslab.sjtu.edu.cn/~fenggang/service-discovery-techrepo.pdf