

Vehicular Ad Hoc Networks and Trajectory-Based Routing

Yanmin Zhu, Yuchen Wu and Bo Li

Abstract With the rapid advance in short-range radio communications for vehicles (e.g., IEEE 802.11p), vehicular ad hoc networks have emerged as a new paradigm of mobile ad hoc networks. Within a vehicle ad hoc network, there are several types of wireless communications, including vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and infrastructure-to-vehicle (I2V). A wide variety of existing applications are being developed based on vehicular ad hoc networks, such as traffic safety, transport efficiency, and entertainment on the move. A key enabling technology for vehicular ad hoc network is message routing among vehicles. Efficient message routing is particularly challenging for vehicular ad hoc networks because of frequent network disruption, fast topological change and mobility uncertainty. The vehicular trajectory knowledge plays a key role in message routing. By extracting mobility patterns from historical vehicular traces, we develop trajectory predictions by using multiple order Markov chains. We then present routing algorithms taking full advantage of predicted probabilistic vehicular trajectories. We carry out extensive simulations based on large datasets of real GPS vehicular traces. The simulation results demonstrate that the trajectory based routing algorithms can achieve higher delivery ratio at lower cost.

1 Introduction

With the rapid advance in short-range radio communications for vehicles (e.g., IEEE 802.11p), vehicular ad hoc networks have emerged as a new paradigm of mobile ad hoc networks. A vehicular ad hoc network is a network of vehicles which communicate with each other via short-range wireless communications [1]. Within a vehicle ad hoc network, there are several types of wireless communications,

Y. Zhu (✉) · Y. Wu · B. Li
Department of Computing Science and Engineering, Shanghai Jiao Tong University,
Shanghai, China
e-mail: yzhu@cs.sjtu.edu.cn

including vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and infrastructure-to-vehicle (I2V). Vehicular ad hoc networks have many appealing applications, such as driving safety [2], intelligent transport [3, 4], and entertainment on the move.

Efficient message routing is of central importance to vehicular ad hoc networks [5–11]. Message routing in a vehicular ad hoc network may introduce nonnegligible delivery latency because of frequent network topology disconnections. Thus, we should stress that the message routing in vehicular ad hoc network is suitable for those applications which can tolerate certain delivery latency. There are many examples of such applications, including peer-to-peer file sharing, entertainment, advertisement, and file downloading.

The knowledge of future vehicular traces plays a key role for optimal routing. Existing routing algorithms heavily rely on prediction of vehicle mobility. However, they have adopted only simple mobility patterns, such as the spatial distribution and inter-meeting time distribution, which support coarse-grained predictions of vehicle movements. Some algorithms [9, 12] assume random mobility in which vehicles move randomly in an open space or a road network. This model is simple but far from the reality. Some other algorithms assume simple mobility patterns such as exponential inter-meeting times and regular spatial distributions. As a result, prediction results based on these simple patterns are of limited value to efficient routing in vehicular ad hoc networks. In addition, many of existing algorithms ignore the fact that links in a vehicular ad hoc network have unique characteristics [11]. On the one hand, a link is typically short-lived. This suggests that the capacity of the link is limited. Thus, the order for forwarding packets becomes important. On the other hand, links in a densely populated area may interfere with each other. This indicates that link scheduling becomes necessary.

To overcome the limitations of existing algorithms, this chapter presents an approach to taking full advantage of trajectory predictions. By analyzing an extensive dataset of real vehicular traces, we find that there is strong spatiotemporal regularity with vehicle mobility. More specifically, our results based on conditional entropy analysis demonstrate that the future trajectory of a vehicle is greatly correlated with its previous trajectory. Thus, we develop multiple order Markov chains for predicting future trajectories of vehicles. With the available future trajectories of vehicles, we present an analytical model and theoretically derive the delivery probability of a packet. We develop an efficient global algorithm and a fully distributed algorithm which needs only localized information. The two algorithms jointly consider packet scheduling and link scheduling. We evaluate the algorithms with extensive trace driven simulations, based on the trace dataset collected in Shanghai and Shenzhen. The results demonstrate that our algorithm considerably outperforms other algorithms in terms of delivery probability and delivery efficiency.

The rest of the chapter is organized as follows. In Sect. 2, we briefly introduce vehicular ad hoc networks. Next, we present the network model and formulate the problem of message routing in vehicular ad hoc networks. In Sect. 4 we analyze vehicle traces and present a mobility model based on Markov chain. Section 5 describes a global routing algorithm and distributed routing algorithm. Simulation results are presented in Sect. 6. The chapter is concluded in Sect. 7.

2 Vehicular Ad Hoc Networks

As illustrated in Fig. 1, a vehicular ad hoc network is simply a system of vehicles that communicate with each other with short-range radios, such as 5.9 GHz DSRC. Besides inter-vehicle communications, vehicles can also communicate with the infrastructure that is comprised by roadside units (RSUs).

Conceptually, inexpensive wireless local area network (WLAN) technologies, such as WiFi, can be used for V2V, V2I, and I2V communications. However, new challenges arise by high vehicle speeds and highly dynamic network environments. In addition, new requirements are imposed by applications of vehicular ad hoc networks, especially by those safety-critical applications, e.g., high reliability of message delivery, low packet latency.

In response to the new challenges and new requirements, new standards have been developed. The American Society for Testing and Materials (ASTM) subcommittee E17.51 examined issues concerning vehicle roadside communications and proposed DSRC. It is largely based on the IEEE 802.11a standard, including the physical (PHY) layer and the media access control (MAC) layer. Currently, the IEEE 802.11p working group continues the main tasks of DSRC.

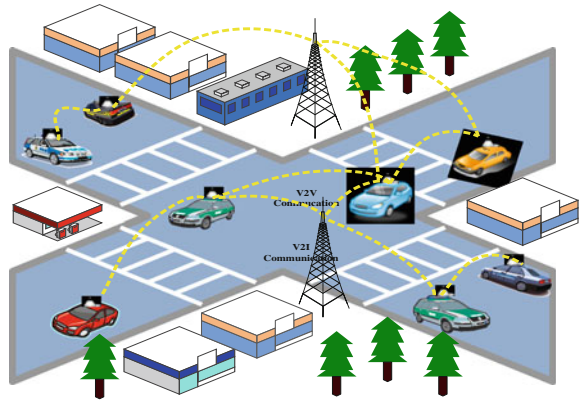
Compared with traditional mobile ad hoc networks, vehicular ad hoc networks face several new challenges.

First, vehicular ad hoc networks are subject to frequent disruptions. It is difficult to find a connected path between a pair of source and destination in vehicular ad hoc networks. This is caused by high mobility and uneven distribution of vehicles over the network.

Second, vehicles usually move at a high speed of up to 80 Km/h. Two vehicles can communicate only when they are within the communication range. Recent study has shown that the contact duration in case of a vehicle and a static access point is as short as 10s on average [13].

Finally, there is a great deal of uncertainty associated with vehicle mobility. Vehicles move at their own wills. It is difficult, if not impossible, to gain the complete knowledge about the vehicular trace of future movement, i.e., the position of the vehicle at a given point in time. For message routing in a vehicular ad hoc network, a relay node must decide how long a packet should be kept and which node a given packet should be forwarded to. Existing study [14, 15] shows that it is possible to find an optimal routing path when the knowledge of future node traces is available, which is NP-hard though. However, it is impractical to have prior knowledge about future traces of nodes.

Fig. 1 The illustration of vehicle-to-vehicle communications and vehicle-to-infrastructure communications in a road intersection



3 Network Modelling and Routing Complexity

3.1 Network Model

The vehicular ad hoc network is modeled as a set of nodes, N . When two nodes, i and j , are in the communication range (denoted by D), i.e., $d_{ij} < D$, there is a link between the two nodes and they can communicate with each other while the link exists.

The position of node n at time τ is denoted by $p_n(\tau)$. The time is slotted. Therefore, the trajectory of node n is a sequence of positions, denoted by

$$T_n = \langle p_n(0), p_n(1), \dots, p_n(\tau) \rangle. \quad (1)$$

The distance between two nodes, i and j , at time τ is denoted by $d_{ij}(\tau)$

Links in the network are changing with the relative positions of the nodes. The set of all possible links at time τ is denoted by $L(\tau)$,

$$L(\tau) = \{l_{i,j} | d_{i,j}(\tau) \leq D; \forall i, j \in N\}. \quad (2)$$

Links are assumed to have the same capacity in theoretical analysis.

The vehicular ad hoc network tries to deliver a set of packets, denoted by Φ , and the packets are of equal size. A packet has a source, $\delta(p)$, and a destination, $\psi(p)$. A packet is copied and forwarded from one node to another if there is a link between them. A packet group, $\theta(p)$, is introduced to denote all the copies of packet p in the network. The size of packet group, $|\theta p|$, increases when there is a new copy-forward process of packet p in the network.

3.2 Formulation of Routing in Vehicular Ad Hoc Networks

The main goal of message routing is to move the packets from their sources to respective destinations. The delivery performance objectives include delivery ratio, delay and efficiency. In the following, we first analyze the properties of individual packets, and then show the objectives in a global view.

The delivery probability is the chance of a packet successfully delivered from its source to its destination. The delivery probability of packet p , denoted by ρ_p , depends on the routing strategy, Υ , and the trajectories of nodes,

$$\rho_p = f_{\Upsilon}(\delta(p), \psi(p), \{T_n | n \in N\}). \quad (3)$$

With given trajectories, ρ_p can be calculated at the beginning and it does not change over time. But in the real world where the future traces are uncertain, this delivery probability can be considered as a random variable.

The delay of a packet, denoted by σ_p , is defined as the total transmission time spent for the network to deliver the packet to its destination. Analysis of delay is similar to that of delivery probability. The delay consists of two parts. One part is the time already spent, and the other is time to spent,

$$\sigma_p(\tau) = \tau + \Gamma_{\Upsilon}(\delta(p), \psi(p), \{T_n\}). \quad (4)$$

When $\sigma_p \tau = \tau$, the delivery of the packet is complete.

The cost of the transmission of a packet, denoted by ζ_p , is defined as the total number of times that a packet is forwarded. A new copy is created when a packet is forwarded. Thus, the total number of copies indicates the delivery cost. The cost is therefore defined as

$$\zeta_p = |\theta(p)|. \quad (5)$$

One of the common objectives of vehicular ad hoc networks is to maximize *delivery ratio*. The delivery ratio is defined as the proportion of successfully delivered packets to the total packets to be transmitted. The number of total packets is often omitted in calculations because it is fixed and not affected by routing strategies. Thus, one objective can be given by

$$\max \sum_{p \in \Phi} \mathbb{E}[\rho_p]. \quad (6)$$

Thus, the problem is to find the optimal routing of the packets through the vehicular ad hoc network that meets (6).

Two other common objectives are minimization of delay and minimization of total cost, which are given respectively by

$$\min \sum_{p \in \Phi} \mathbb{E}[\sigma_p], \text{ and } \min \sum_{p \in \Phi} \mathbb{E}[\zeta_p]. \quad (7)$$

The efficiency is defined as the ratio of total successfully delivered packets to the total cost, given by

$$\ell = \sum_{p \in \Phi} \mathbb{E}[\rho_p] / \sum_{p \in \Phi} \mathbb{E}[c_p]. \quad (8)$$

A high efficiency indicates that one algorithm achieves high delivery ratio at a low cost.

3.3 Analysis of Routing Complexity

The optimal routing for achieving the objectives mentioned above is extremely difficult to design. For analysis simplification, we first assume that the routing algorithm have the complete knowledge of all vehicles and their future traces, and the set of packets to be transmitted. Note that this is impractical in the real world.

Without loss of generality, we take the objective of maximizing the delivery ratio for example. Objective (6) can be written as,

$$\max \sum_{p \in \Phi} \mathbb{E}[f_{\Upsilon}(\delta(p), \psi(p), \{T_n | n \in N\})]. \quad (9)$$

Theorem 1: This routing problem with objective (9) when the vehicular traces and the set of packets to transmit are given is NP-hard.

The concise proof for the theorem is as follows. The well studied edge-disjoint path (EDP) problem is known to be NP-hard, which tries to find the maximal number of edge-disjoint paths. By reducing the EDP problem to the ideal routing problem, we can prove the ideal routing problem is also NP-hard [8]. The basic reduction procedure is as follows. The vertices of EDP are mapped to the vehicles, and each edge is mapped to a contact of vehicles. The source destination pairs are mapped to the source and destination pairs of the packets to transfer. As a result, the routes are valid edge-disjoint paths.

In practice, the complete knowledge is hard to be obtained because the future knowledge of vehicular traces is usually unavailable. Following this practical constraint, the trajectory of a node is divided into two parts: the past and the future, denoted by

$$\{T_n\} = \{T_n | \tau \leq t\} \cup \{T'_n | \tau > t\}. \quad (10)$$

Let $\lambda_t(p)$ to denote the current position of the packet. The objective at time t for routing becomes

$$\max \sum_{p \in \Phi} \mathbb{E}[\rho_p] = \max \sum_{p \in \Phi} \mathbb{E}[f_{\Upsilon}(\lambda_t(p), \psi(p), \{T_n\}, \{T'_n\})]. \quad (11)$$

The second part of the trajectory is different from the first part. The first part has been fixed while the second part is not fixed and unknown at the time of being.

Since the objective of delay minimization has similar formation, we generalize (11) to the following,

$$\max \sum_{p \in \Phi, n \in N} \Theta(\lambda_t(p), \psi(p), \{T_n |_{\tau \leq t}\}, \{T'_n\}). \quad (12)$$

where Θ denotes the global metric for the whole network. If we omit the inputs, the objective becomes

$$\max \sum_{p \in \Phi} \Theta_{p, N, t}. \quad (13)$$

This is the practical objective with the limited knowledge beyond time t .

4 Mobility Analysis and Modelling

In this section, we quantitatively reveal the spatiotemporal regularity with vehicle mobility. The quantitative analysis is based on mining the large dataset of vehicular traces of more than 4,000 taxis in Shanghai, China, which have a duration of more than two years. The GPS trace of a taxi, equipped with a GPS receiver, was collected by the vehicle periodically sending its instant position to a data collection center at an interval from 10s to several minutes. The taxis operates in the whole urban area of Shanghai, which covers an area of over 120km².

For simplification of discussion, the whole space is divided into Q small grids., and is denoted by S ,

$$S = \{s_0, s_1, \dots, s_{Q-1} | s_i \cap s_j = \emptyset\}. \quad (14)$$

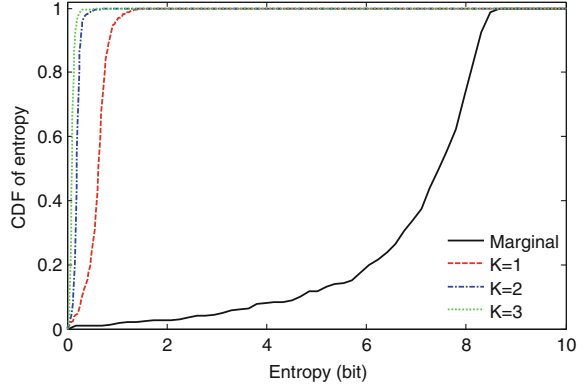
where Q is the total number of grids. The time is slotted. The location of a vehicle at a given time is considered as a random variable which takes state values from the grid space. Let S_i denote the random variable for vehicle i . We reveal the spatiotemporal regularity by computing the marginal and the conditional entropies of S_i given the previous K states.

For vehicle i , suppose we have observed its states for L time slots. The state sequence of the vehicle can be denoted by a vector $T_i = \langle s_0, s_1, \dots, s_{L-1} \rangle$, where $s_j \in S$, $0 \leq j \leq L-1$ is the positional state of vehicle i at time slot j . Suppose that s_j appears o_j times in the vector of T_i , $0 \leq j \leq L-1$. Thus, the probability of vehicle i taking state s_j can be computed as o_j/L . Then, the marginal entropy of S_i is

$$H(S_i) = \sum_{j=0}^{Q-1} (o_j/L) \times \log_2 \frac{1}{o_j/L}. \quad (15)$$

Next, we compute the conditional entropy of S_i given its immediately previous state S_i^1 which has the same distribution with S_i . The conditional entropy is,

Fig. 2 The CDFs of marginal entropy and conditional entropies of a vehicle positional state. The grid size is $200\text{ m} \times 200\text{ m}$, and the number of vehicles is 500



$$H(S_i|S_i^1) = H(S_i, S_i^1) - H(S_i). \quad (16)$$

To derive the conditional entropy, we have to derive the entropy of the joint random variable (S_i, S_i^1) . By using the state sequence T_i , we derive a sequence of 2-tuples, $T_i^1 = \langle (s_0, s_1), (s_1, s_2), \dots, (s_{L-2}, s_{L-1}) \rangle$. By counting the occurrences of (s_{j-1}, s_j) , denoted by $o_{j-1,j}$, we can get its probability. Thus, the joint entropy of the two dimensional random variable of (S_i, S_i^1) is,

$$H(S_i, S_i^1) = \sum_{j=1}^{Q-1} \frac{o_{j-1,j}}{L-1} \times \log_2 \frac{1}{o_{j-1,j}/(L-1)}. \quad (17)$$

By generalizing the previous computation, we can compute the conditional entropy of S_i given its immediately previous K states $S_i^1, S_i^2, \dots, S_i^K$

$$H(S_i|S_i^1, S_i^2, \dots, S_i^K). \quad (18)$$

In essence, by showing the marginal entropy we reveal spatial regularity, which characterizes the uncertainty of a vehicle residing a location in the space. By showing the conditional entropy given the previous states, we reveal the spatiotemporal regularity of vehicle mobility. This characterizes the uncertainty of a vehicle residing a specific location when its previous states are given. This demonstrates the correlation of a vehicle's positional states over different times.

In Fig. 2, the cumulative distribution functions (CDFs) of the marginal and the conditional entropies for are shown. The conditional entropies $K = 1, 2, 3$ are significantly smaller than the marginal entropy. This implies that the uncertainty of the positional state becomes smaller when the previous states are known. We also find that when K becomes larger, the entropy continues to decrease. This suggests that more previous states help further reduce uncertainty. However, the improvement quickly stalls as K increases. This gives the guidance to the order selection for trajectory prediction using multiple order Markov chains.

4.1 Modelling Vehicular Mobility with Markov Chain

A mobility pattern of a node characterizes the regularity of its mobility.

Definition 1. The mobility pattern of node n , denoted by MP_n , is defined as a pair of random variables, $MP_n = (\alpha, \beta)$. The probability distribution of the mobility pattern characterizes the regularity of the vehicle's mobility.

We develop the following mobility pattern for characterizing the spatiotemporal regularity of vehicle mobility. For vehicle i , its mobility pattern is defined as $M_i = (H, F)$, where H is a vector of past positional states, $H = \langle H_0, H_1, \dots, H_{K-1} \rangle$, $K \geq 1$, and F is the future positional state. The mobility pattern characterizes the frequencies of F following H in the vehicle's traces by computing the probability distribution of the two-dimensional random variable M_i . This can be achieved by analyzing the historical traces of the vehicle, as discussed in Sect. 3.

The notion of mobility pattern is general enough to cover the regularities that have been studied. For the study of inter-meeting times, the mobility pattern is

$$MP_n = (A, B), \quad (19)$$

where A is a random variable representing the inter-meeting time of vehicle n and a second vehicle, and B is the random variable denoting the second vehicle. Based on this mobility pattern, we are able to describe the distribution of inter-meeting times between n and any other vehicle.

For the study of spatial distribution of a vehicle, the mobility pattern is

$$MP_n = (S, \emptyset), \quad (20)$$

where S is the random variable representing the spatial state and \emptyset denotes a null random variable.

To predict the trajectory of a vehicle, existing simple patterns are inadequate. In the next subsection, we present our method of trajectory prediction by developing multiple order Markov chains based on the spatiotemporal mobility pattern.

4.2 Predicting Future Trajectories

The problem of predicting the trajectory of vehicle i at time t is to compute the future trajectory $T_i|_{\tau > t}$, given the trajectory before t , $T_i|_{\tau < t}$. A trajectory, T , can be described by a sequence of positional states,

$$T = \langle s^0, s^1, \dots, s^\tau, \dots \rangle. \quad (21)$$

Because of the uncertainty associated with vehicle mobility, there may exist a number of possible future trajectories.

Definition 2. The set of all possible trajectories of node n is defined as a trajectory bundle, denoted by TB_n , which can be characterized by

$$TB_n = \langle D^n(1), D^n(2), \dots, D^n(\tau), \dots \rangle. \quad (22)$$

where $D^n(\tau)$ is the probability distribution of spatial states at future time τ .

$$D^n(\tau) : P_s^n(\tau) | s \in S. \quad (23)$$

where $P_s^n(\tau)$ is the probability of the node n appearing in state s at time τ .

To predict the trajectory of a vehicle, it is essentially to calculate the trajectory bundle of the vehicle. For trajectory prediction, we develop multiple order Markov chains for predictions. The key is to establish the matrix of transition probabilities. The transition probabilities are computed on an individual vehicle's basis, since different vehicle possesses different regularities.

For a vehicle, we can easily derive its transition matrix (denoted by X) from its mobility pattern. When we are using a K -order Markov chain, an element, $x_{ij} \in X$, represents the transition probability from H_i to F_j , where H_i is a sequence of positional states, $H_i = \langle h_0^i, h_1^i, \dots, h_{k-1}^i \rangle$, and F_j is a single state. Then, $x_{ij} = Pr(H_i, F_j)$, where $Pr(\cdot)$ is the probability distribution function of the mobility pattern of the vehicle.

By applying the K -order Markov chain, we can compute the trajectory bundle as follows. Given the current trajectory of node n is $T_c^n = \langle s_{-k+1}, s_{-k+2}, \dots, s_0 \rangle$, the initial distributions for T_c^n are

$$D^n(\tau) : \begin{cases} P_{s_\tau}^n(\tau) = 1 \\ P_s^n(\tau) = 0, s \neq s_\tau \end{cases}, \quad (\tau \leq 0). \quad (24)$$

The distributions of spatial state at future times can be iteratively calculated. For a single trajectory (denoted by $\langle s_1, \dots, s_k \rangle, s$), its probability is

$$P_{\langle s_1, \dots, s_k \rangle, s}^n(\tau) = x_{\langle s_1, \dots, s_k \rangle, s} \times \prod_{i=1}^K P_{s_i}^n(\tau - K + i). \quad (25)$$

$P_{si}^n(t)$ are derived from the previous distributions $x_{\langle s_1, \dots, s_k \rangle, s}$ and is defined in the transition matrix. Then, $D_n(\tau)$ can be derived by

$$D^n(\tau) : P_s^n(\tau) = \sum_{all \langle s_i \rangle} P_{\langle s_i \rangle, s}^n(\tau), \quad \tau > 0. \quad (26)$$

4.3 Computing Inter-Vehicle Delivery Probability

With the predicted trajectories, we are able to derive the encounter probabilities which are required for computing the eventual delivery probabilities.

Given the trajectory bundles of node i and j , $D^i(\tau)$ and $D^j(\tau)$ are known. Let $\varepsilon_{i,j}(\tau)$ denote the encounter probability of the two nodes at time τ . It can be calculated by

$$\varepsilon_{i,j}(\tau) = \sum_{s \in S} P_s^i(\tau) \times P_s^j(\tau). \quad (27)$$

Then, the encounter probability, $\varepsilon_{i,j}$ is given by

$$\varepsilon_{i,j} = 1 - \prod_{\tau=t}^T (1 - \varepsilon_{i,j}(\tau)), \quad (28)$$

where T denotes the max prediction range of time.

If the overall objective is to minimize the delivery delay, the estimated encounter time for the two nodes can be derived by

$$\eta_{i,j} = \frac{\sum_{\tau=t}^T \tau \times \varepsilon_{i,j}(\tau)}{\sum_{\tau=t}^T \varepsilon_{i,j}(\tau)}. \quad (29)$$

Therefore, the estimated delay of a packet can be calculated through trajectories in a similar way with delivery probability computation.

5 Trajectory-Based Routing

In this section we first present the global routing algorithm and the distributed routing algorithm.

5.1 Global Algorithm

Designing the optimal routing for achieving (13) is difficult, especially without any knowledge about future vehicular traces. We present a global routing algorithm with only predicted trajectories of vehicles.

The global metric of a packet changes after intermediate transfers over time. Consider that a packet, p , is forwarded from node i to j at time t . After the forwarding, the increment in the global metric becomes

$$\Delta \Theta_{p,N,t} = \Theta_{p,N,t}(j) - \Theta_{p,N,t}(i). \quad (30)$$

And the original objective becomes

$$\max \sum_{p \in \Phi} \left(\sum_{\tau=0}^{t-1} \Delta \Theta_{p,N,\tau} + \Delta \Theta_{p,N,t} \right). \quad (31)$$

Since the sum of increments before time t has been fixed, it can further be written as

$$\max \sum_{p \in \Phi} \Delta \Theta_{p,N,t}. \quad (32)$$

Item $\Delta \Theta$ can be considered as the current metric for packet p .

At time t , there may exist a number of links that may inference with each other, and it is impossible for all the links to be active simultaneously. The routing algorithm must schedule the links, i.e., choose a subset of links from all possible links to maximize the sum of current metrics. Thus, the objective becomes

$$\max \sum_{p \in \Phi, l \in L(t)} (\Delta \Theta_{p,N,t} \times l_{i,j}). \quad (33)$$

The optimal routing performs two tasks, i.e., link scheduling and packet scheduling. Before the routing decision can be made, we have to obtain the metric increment of every packet over all possible transfers. When all the increments are available, the optimal routing chooses such a set of links and a set of packet transfers that meet (33).

We should emphasize that even when all $\Delta \Theta_{p,N,t}$ are available, finding the best links and packet transfers is still an NP-hard problem. A brief proof is as follows. This problem can be considered as a weighted maximum independent set problem that has been proved to be NP-hard. To solve this problem, a number of existing heuristic algorithms [16] can be used, in which $\Delta \Theta_{p,N,t}$ are considered as link weights.

As mentioned before, we take delivery probability maximization as example. The delivery probability of packet, ρ_p , depends on the encounter probability of every two nodes, denoted by $\epsilon_{i,j}, \forall i, j \in N$. In the following we derive the relationship between metric increments and encounter probability.

The probability for the packet to be delivered in no more than one hop is

$$\rho_p^1 = \epsilon_{\lambda(p), \psi(p)}. \quad (34)$$

Then, for a two hop delivery with the two-hop route of $\langle \lambda(p), n_1, \psi(p) \rangle$, the probability is

$$\rho_p^{\langle \lambda(p), n_1, \psi(p) \rangle} = \epsilon_{\lambda(p), n_1} \times \epsilon_{n_1, \psi(p)}. \quad (35)$$

Thus, the probability for packet being delivered in two hops is

$$\begin{aligned} \rho_p^2 &= 1 - \prod_{n \in N, n \neq \lambda(p), \psi(p)} (1 - \rho_p^{\langle \lambda(p), n, \psi(p) \rangle}) \\ &= 1 - \prod_{n \in N, n \neq \lambda(p), \psi(p)} (1 - \epsilon_{\lambda(p), n} \times \epsilon_{n, \psi(p)}). \end{aligned} \quad (36)$$

The delivery probability of the packet with the h -hop route $\langle \lambda(p), n_1, \dots, n_{h-1}, \psi(p) \rangle$ is

$$\rho_p^{\langle \lambda(p), n_1, \dots, n_{h-1}, \psi(p) \rangle} = \varepsilon_{\lambda(p), n} \times \prod_{i=1}^{h-2} \varepsilon_{n_i, n_{i+1}} \times \varepsilon_{n_{h-1}, \psi(p)}. \quad (37)$$

Then the set of all h -hop routes (denoted by \mathcal{E}) which starts from λ and ends at ψ is denoted by

$$\mathcal{E}^h(\lambda, \psi) = \{ \langle \lambda, n_1, \dots, n_{h-1}, \psi \rangle \mid n_i \in N; n_i \neq \lambda, \psi \}. \quad (38)$$

Thus, the h -hop delivery probability is calculated by

$$\rho_p^h = 1 - \prod_{r \in \mathcal{E}^h(\lambda(p), \psi(p))} (1 - \rho_p^r). \quad (39)$$

Therefore, the total delivery probability is

$$\rho_p = 1 - \prod_{0 < h < H} 1 - \rho_p^h. \quad (40)$$

Constant H acts as the hop limit.

From the previous analysis, we can find that encounter probability is the key to computing the overall delivery probability of a packet. Since it is impossible to know future movements of vehicles, the knowledge of encounter probability of each pair of vehicles is not immediately available.

Fortunately, we have observed that there is strong spatiotemporal regularity with vehicle mobility. Based on this observation, we present mobility patterns to characterize this regularity. With the mobility pattern, we are enabled to predict the trajectory of a vehicle. With trajectories of vehicles, we can effectively compute the encounter probability of two vehicles.

We should stress that the encounter probability computed based on this method is more accurate than those derived from simple patterns, such as inter-meeting times and spatial distribution. Our prediction effectively makes use of both the current state information and the historical information of vehicle movement. In addition, the historical information is purely based on individual vehicles. This overcomes the problem with traditional methods that require historical information about any pair of vehicles, which introduces additional overhead.

The global algorithm is described as follows. According to Sect. 4.2, each vehicle which has a transition matrix obtained from historical data can calculate its future trajectory bundle using its current position. With the trajectory bundles, the encounter probability of every two vehicles is known from (27). Then, the delivery probability of each packet in the network through each possible path can be calculated by (28) and (40). With these probabilities, the global algorithm solves an optimization problem to determine which the best next hop for each packet is. The algorithm repeatedly schedules the delivery until all the packets are delivered.

5.2 Distributed Algorithm

The global algorithm requires the complete knowledge of all packets and vehicles. More specifically, for each packet, the knowledge of its source and destination, and the current position must be known. For each vehicle, the knowledge of its position and the mobility pattern must be known. These pieces of knowledge are not available in a distributed setting. Thus, we design a practical, distributed algorithm with which a vehicle requires only limited and localized knowledge.

The distributed algorithm consists of two fundamental building blocks. In the first block, the new objective for each individual vehicle is designed, since it is impractical for individual vehicles to compute the global objective. In the second block, we define the metadata for vehicles to exchange with each other at meetings.

In the distributed setting, the knowledge of nodes and packets is usually incomplete. Therefore, a global optimization (32) is hard for it is impossible to compute the global metric Θ and the current metric $\Delta\Theta$ (defined in (11), (12), and (30)). In this case, we have to design a new metric based on which an individual vehicle makes routing decisions. Let the incomplete set of nodes and packets be denoted by Φ and Φ' , respectively. Then, the new objective becomes

$$\max \sum_{p \in \Phi'} \Delta\Theta_{p,N',t}. \quad (41)$$

$\Delta\Theta_{p,N',t}$ becomes the new metric of packet p at time t , which is local and can be computed by individual vehicles for each packet. When making routing decisions, a node maximizes the sum of local metrics of all the packets it knows.

It would be better for a node to have more knowledge about nodes and packets. Since it is difficult for a node to have the complete knowledge, we design a distributed protocol for the nodes to exchanging information when they meet each other. By this way, the knowledge can be propagated throughout the network.

For exchanging information, we define metadata, which include two parts of information. The first part is about the mobility pattern and the most update position of each vehicle (time stamped). The second part is about the packets that the node carries. Note that for a relatively stable set of vehicles, the mobility pattern reflects the regularity of a vehicle's mobility. Thus, it is relatively stable and therefore it is no need to update the mobility patterns frequently.

As a distributed algorithm, each vehicle executes the routing algorithm independently. For each vehicle, a routine procedure is invoked each time it finds a new communication neighbor. For neighbor discovery, it is required that every vehicle periodically broadcasts hello messages so that other vehicles can discover it when it enters their communication ranges.

In the following we describe the procedure, supposing that vehicle n finds another vehicle, n' , entering its communication range. The Pseudo code description of this procedure is shown in Fig. 3.

First, node n updates its neighbor set (Λ_n). Then, node n and n' exchange their metadata. The metadata of a node, i , denoted by \mathcal{M}_i include two metadata sets, \mathcal{P}_i

Procedure 1 ($n, n', \Lambda_n, \mathcal{M}_n, \mathcal{M}_{n'}, t$)
Variables n : the node; n' : new neighbor; $\mathcal{M}_n, \mathcal{M}_{n'}$: metadata; Φ_n : packets of node n ; Λ_n : neighbors of n ; t : current time; N : total nodes

1. $\Lambda_n = \Lambda_n \cup \{n'\}$,
 2. n exchanges metadata with n' : $n \leftarrow \mathcal{M}_{n'}$, $n' \leftarrow \mathcal{M}_n$
 3. n calculates $\Delta\Phi_{n'} = \Phi_n - \Phi_{n'}$ from $\mathcal{M}_n, \mathcal{M}_{n'}$
 4. For every packet $p \in \Delta\Phi_{n'}$, calculate metric $\Delta\theta_{p,N,t}(n') = \rho_p n' - \rho_p(n)$
 5. Sort all packets $p \in \Delta\Phi_i$ according to metric $\Delta\theta_{p,N,t}(i)$ for all $i \in \Lambda_n$ in the decreasing order
 6. Transmit the packets in this order
-

Fig. 3 Pseudo code of the procedure of the distributed algorithm, which is executed upon each contact with a vehicle

and \mathcal{V}_i . Set \mathcal{P}_i contains the metadata about the packets that node carries, including identification and source-destination pair. Set \mathcal{V}_i contains the metadata about the vehicles known to node i , including ID, most update position and mobility pattern.

Next, it recalculates the metrics for all the packets it carries, and sort the packets respect to the new metrics in the decreasing order. The packets will then be transferred in the sorted order. Note that if the node has already been transferring a packet, the transmission of this packet is not interrupted. After its completion, the packets shall be transmitted according to the new order. By ordering the packets, we are essentially doing the link scheduling in a distributed way. However, before a packet transfer can be started, a vehicle has to follow media access control to avoid potential collisions.

As packets are transferred between vehicles, the packet set, \mathcal{P}_i , is updated whenever the vehicle receives a new packet from its neighbors.

6 Simulation Study

We evaluate our algorithms with the performance metric of delivery ratio, delay, cost and efficiency. These metrics have been defined in Sect. 2. We compare our algorithms with several other algorithms, which is to be introduced shortly.

The simulations are conducted with the three datasets of real GPS vehicular traces. One dataset includes the vehicular traces of more than 4,000 taxis collected in Shanghai, the second dataset is the traces of more than 2,000 buses in Shanghai and the third dataset consists of vehicular traces of more than 12,000 taxis collected in Shenzhen. The Shanghai dataset covers a duration of two years and the Shenzhen dataset a duration of one month. The whole urban area of Shanghai is 133 km in length and 69 km in width, and that of Shenzhen is about 27 km in length and 97 km in width. Table 1 shows the summary of the three trace datasets.

The whole space is divided into grids, and the grid size is three kilometers by default and will be varied in the impact study of grid size. The communication range

Table 1 Summary of trace datasets

Trace	Shanghai taxi	Shanghai bus	Shenzhen taxi
# of nodes	4,000	2,000	12,000
Contacts (per hour)	352	891	425
Granularity (second)	60	60	60
Duration (hour)	17,280	17,280	720
Range (km \times km)	133 \times 69	133 \times 69	27 \times 97

is 300 m. We consider link interfaces, and each node can communicate with one neighbor at any time. We randomly select a subset of 400 taxis or buses from the complete trace for simulations.

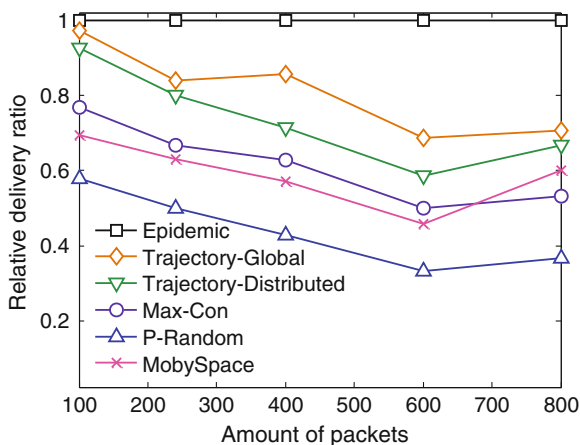
For each packet, we randomly select its source and destination. The packets are injected at different times. Every packet has the same size and priority. The maximum hop and the maximum time-to-live (TTL) of each packet is set to 20–2 h, respectively. The number of packets is varied from 100 to 800 to study different loads of the network.

The order of Markov chains K is set to two. From conditional entropy analysis, we have already found that a higher order beyond two gives little reduction in uncertainty. To build the mobility pattern for each vehicle, its vehicular trace of the past 15 days is used. Each data point is the average over five different runs of trace-driven simulations.

We compare our routing algorithms with the following algorithms

- **Flooding** [17], also known as epidemic routing, is a simple algorithm. Each node forwards all the packets it carries to any node it meets. This algorithm provides an upper bound on delivery ratio and a lower bound on delivery delay. It introduces very high cost, which is the major defect.
- **P-Random** [15] is an opportunistic routing algorithm, which randomly decides whether to forward a packet to another node. In simulations, the probability is set to 0.4. This algorithm represents the algorithms without predictions.
- **MobySpace** [9] is a routing algorithm for delay tolerant networks. It obtains the probability of each node residing in each possible location by analyzing the historical trace data. It assumes that every node has the full knowledge of such probabilities of all vehicles. It then computes the contact probability of every two vehicles based on the probability distribution, and then routing decisions can be made.
- **Max-Contribution** [11] is a routing algorithm using a simple mobility pattern that only considers the inter-meeting times of exponential distribution. This pattern makes prediction independent of the current location of a vehicle.

Fig. 4 Delivery ratio versus amount of packets, with Shanghai taxi dataset



6.1 Performance Results

We present performance results of our algorithms compared against other algorithms. First, we show the results using the Shanghai dataset of real taxi traces. To characterize the mobility pattern, the vehicular trace of the recent 15 days is used.

We vary the amount of packets and compare the six algorithms in terms of delivery ratio, average delay and total cost under different loads of network. For a given setting of amount of packets, the same set of packets and the set of nodes are used for all the six algorithms.

Since the delivery ratio is largely affected by the simulated time length, we use the metric of relative delivery ratio instead of bare delivery ratio, which is the delivery ratio of each algorithm normalized by that of Flooding. In Fig. 4, the performance of the six algorithms in terms of relative delivery ratio is shown. We can see that our algorithms perform better than P-Random, MobySpace and Max-contribution. Among the six algorithms, Flooding performs the best, as expected. In general, the algorithms that use predictions produce better delivery ratios than the algorithms that make no predictions. Our algorithms are better than MobySpace and Max-Contribution because our algorithms use not only the historical traces information but also the current state and the previous states. Predicted trajectories of vehicles help to find better routing paths. The global algorithm is better than the distributed one since it has the global, complete network information. We can also find that when the amount of packets increases, the overall delivery ratios of all algorithms decrease. The reason is that the overall capacity of the network is limited. By injecting more packets, the packets may compete for network resources and as a result, fewer packets can be delivered in the end.

In Fig. 5, average delay against amount of packets is plotted. Our algorithms have a lower delay than P-Random, MobySpace and Max-Contribution. As expected, Flooding has the smallest delay. The average delays of our algorithms are slightly

Fig. 5 Average delay versus amount of packets, with Shanghai taxi dataset

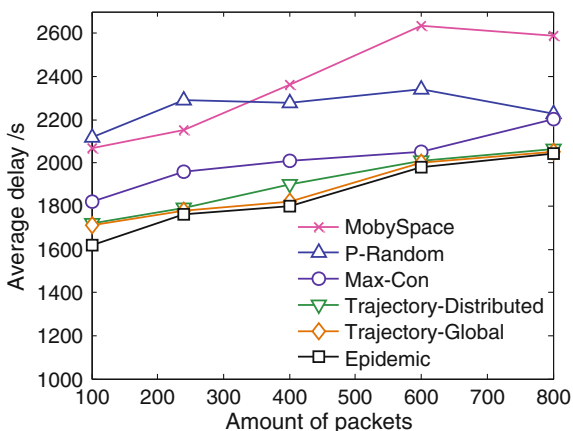
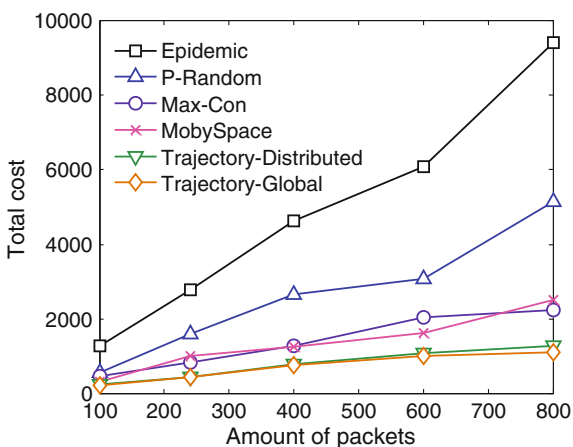


Fig. 6 Total cost versus amount of packets, with Shanghai taxi dataset

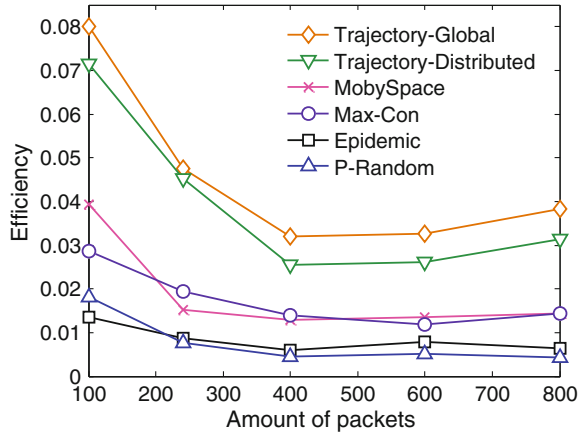


larger than that of flooding. This performance gain of our algorithms is mainly due to the fact that routing paths with high delivery probabilities usually lead to shorter delays. Since our algorithms can select routing paths of high delivery probabilities, the resultant delay is low.

Figure 6 shows the costs of the six algorithms. We can find that our algorithms have lower costs, better than all the other algorithms. The main reason is that by effectively predicting the trajectories of vehicles, we only consider the paths that lead to eventual delivery with high probability. Therefore, unnecessary packet transfers are greatly reduced.

In Fig. 7, we compare the efficiencies of the six algorithms. We can see that our algorithms have higher efficiency, better than all the other algorithms. Flooding and P-Random have a similar efficiency and are much worse than the rest four. This is due to the blindness of Flooding and P-Random when they are making transfer decisions. Max-Contribution and MobySpace have larger efficiency than Flooding

Fig. 7 Efficiency versus amount of packets, with Shanghai taxi dataset



and P-random, but have lower efficiency than our algorithms, since Max-Contribution and MobySpace merely uses a simple mobility pattern of inter-meeting time.

6.2 Impact of Grid Size

We study the impact of grid size. According to the design of the algorithms, two vehicles residing in the same grid are considered to encounter with each other, but they may be unable to communicate in reality. Meanwhile, a larger grid size results in a lower algorithm complexity. The simulation is performed with the Shanghai taxi dataset with the Shanghai bus dataset.

In Fig. 8, the delivery ratios of the two algorithms are shown for different grid sizes. We find that when the grid size increases, the relative delivery ratio of the two algorithms increases almost linearly. This confirms that a larger grid size makes the algorithms more aggressive, thus injecting more packets into the network. As a result, the delivery ratio becomes higher.

In Fig. 9, we show the performance of delivery delay of the two algorithms when the grid size is varied from 500 to 6,000. We can see that when the grid size increases, the average delays of both the algorithms slightly decrease Shanghai Bus dataset, but increase for Shanghai taxi dataset. Buses have much stronger regularity than taxis. For Shanghai bus dataset, a larger grid size leads to a more aggressive forwarding strategy and helps find shorter routing paths. For Shanghai taxi dataset, however, a more aggressive forwarding strategy size does not lead to short routing paths. The main reason is that predictions for taxis are less effective than for buses and the more aggressive forwarding results in less effective utilization of the limited forwarding opportunities in vehicular ad hoc networks.

In Fig. 10, we look at the performance of total cost of the two algorithms when the grid size varies. It is apparent from the result that the costs of the two algorithms

Fig. 8 Delivery ratio versus grid size

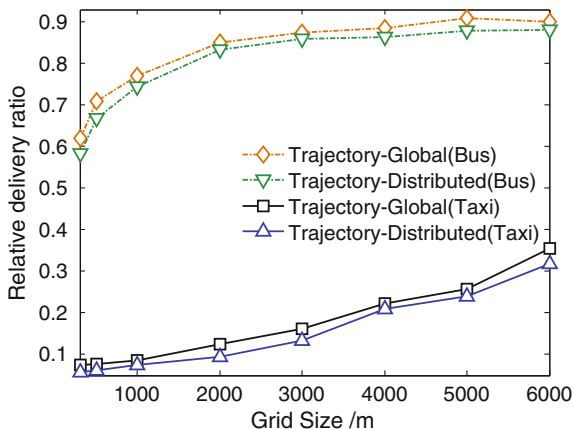
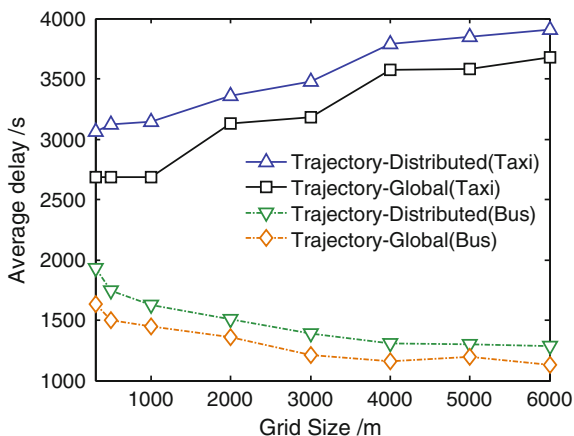


Fig. 9 Average delay versus grid size



quickly increase as the grid size becomes larger. This is easy to understand since with a larger grid size, the algorithms intend to forward more packets and thus inject more packets into the network. Thus, a higher cost is observed when the grid size is larger.

The performance of efficiency for the two algorithms for different grid sizes is shown in Fig. 11. The performance results match our expectation that the efficiency of the algorithms quickly decreases as the grid size becomes larger. This results from the fact that a larger grid size allows the algorithms to make more aggressive data forwarding.

In summary, the grid size controls the tradeoff between delivery ratio and cost (and also efficiency). And importantly, although a grid size introduces higher cost, the delivery delay may not be decreased.

Fig. 10 Total cost versus grid size

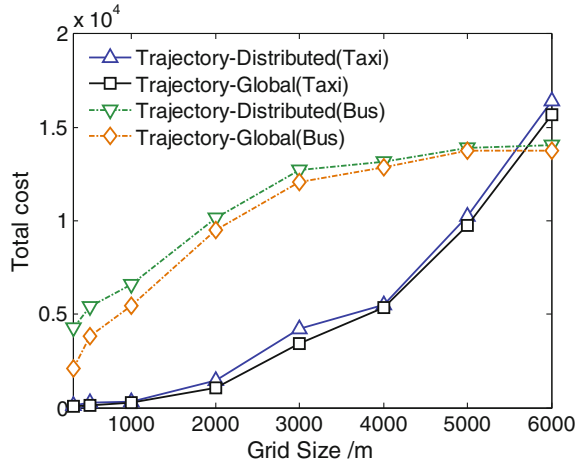
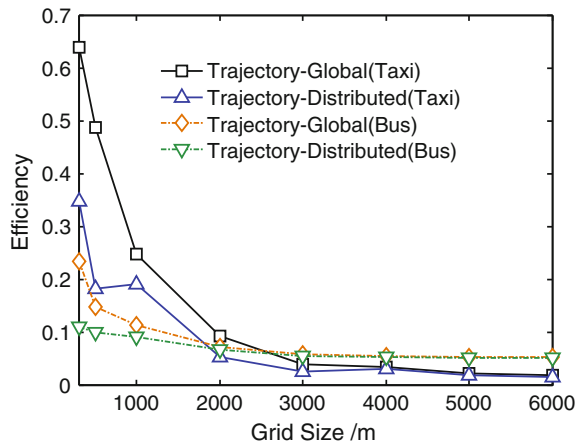


Fig. 11 Efficiency versus grid size



6.3 Impact of Historical Traces

The impact of the use of historical traces is studied. To this end, we vary the length of the historical trace that is used for building the mobility pattern of a given vehicular node. The length is varied from 2 to 16 days. To offset the effect of grid size, we study the performance under two configurations of grid size, 500 and 2,000m. The simulation is performed with the Shanghai taxi dataset.

In Fig. 12, the performance of relative delivery ration of the trajectory based algorithms is shown for different lengths of historical trace. We can see that the relative delivery ratio increases quickly when a longer length of historical trace is used for both the global and the distributed algorithms. This strongly demonstrates that a longer trace builds a better mobility model and is beneficial to prediction of

Fig. 12 Delivery ratio versus length of historical trace

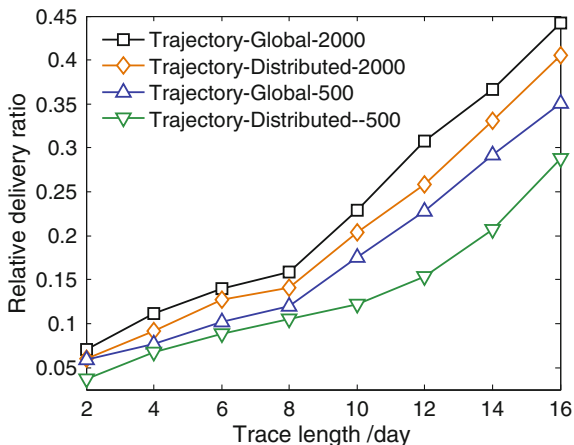
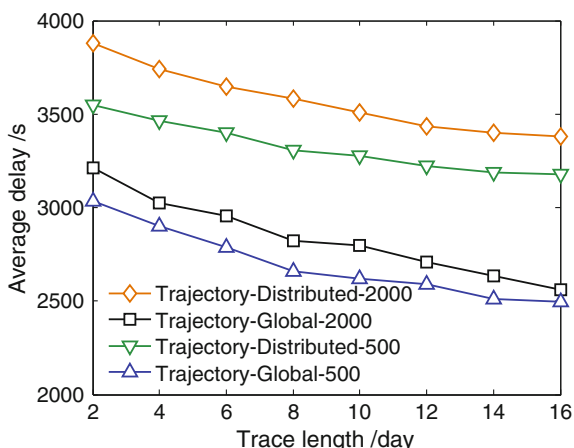


Fig. 13 Average delay versus length of historical trace



vehicular trajectory. This makes the routing process more effective by making use of the predicted trajectory. In addition, we find that when the grid size is 2,000 m, the delivery ratio performance becomes better, as observed previously in the study on the impact of grid size.

In Fig. 13, we show the performance of average delay of the two algorithms for different lengths of vehicular trace. We find that when the length of vehicular trace increases, the average delay drops slightly for both the algorithms. This shows that a longer historical trace is also beneficial to reduction of delivery delay.

Figure 14 shows the performance of the two algorithms in terms of total cost for different lengths of historical vehicular trace. The results show that the cost for each of the two algorithms also increases when the length of historical trace becomes larger. However, it should be noted that the increase in cost when the grid size is 500 m is much slower than the one when the grid size is 2,000 m. Thus, we can

Fig. 14 Total cost versus length of historical trace

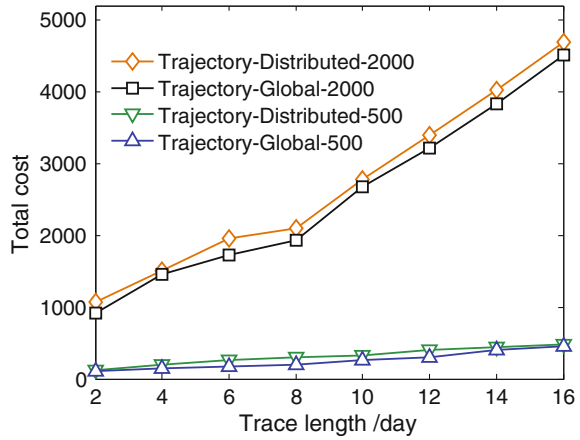
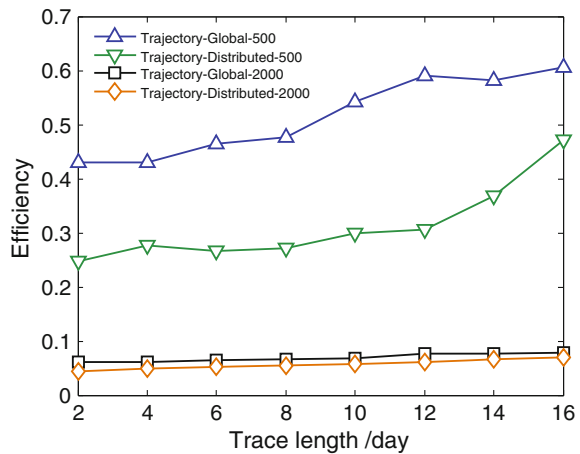


Fig. 15 Efficiency versus length of historical trace



conclude that when a small grid size, e.g., 500 m, is used, a longer historical trace only introduces a modest increase in cost. But when a large grid size is used, the cost rises quickly as the length of historical trace increases.

In Fig. 15, we show the performance of efficiency of the two algorithms for different lengths of historical trace. We can see that as the length of historical trace increases from 2 to 16 days, the efficiency of each algorithm slowly increases. In addition, the efficiency with a grid size of 500 m is much higher than that with a grid size of 2,000 m. And, the efficiency with a grid size of 2,000 m almost does not change as the length of historical trace increases.

In summary, a longer historical trace helps improve the performance of the trajectory based algorithms.

7 Conclusion

Vehicular ad hoc networks are a new type of mobile ad hoc networks. Efficient message routing is of significant importance to applications of vehicular ad hoc networks. However, different from traditional mobile ad hoc networks, existing routing algorithms for mobile ad hoc network cannot directly be used. Although routing of vehicular ad hoc networks has been studied, existing work either assumes the availability of future vehicular traces, e.g., through navigation systems, or fails to make effective use of the vehicular traces. By developing multiple order Markov chains, we predict future vehicle trajectories. The proposed trajectory based algorithms take full advantage of predicted probabilistic vehicle trajectories. Performance results verify that our algorithm outperforms other algorithms. This demonstrates that predicted trajectories do help improve message routing performance in vehicular ad hoc networks.

References

1. Bai F, Stancil, D.D., Krishnan, H.: Toward understanding characteristics of dedicated short range communications (DSRC) from a perspective of vehicular network engineers. Presented at the ACM MOBICOM, (2010)
2. Chisalita, L. Shahmehri, N.: A peer-to-peer approach to vehicular communication for the support of traffic safety applications. Presented at the The 5th IEEE Conference on Intelligent Transportation Systems (2002)
3. Li, Z., Zhu, Y., Zhu, H., Li, M.: Compressive sensing approach to urban traffic sensing. Presented at the IEEE ICDCS (2011)
4. Zhu, H., Zhu, Y., Li, M., Ni, L.M.: SEER: metropolitan-scale traffic perception based on lossy sensory data. In: Proceeding of the IEEE INFOCOM (2009)
5. Leontiadis, I., Mascolo, C.: Geopps: geographical opportunistic routing for vehicular networks. In: Proceedings of the IEEE WoWMoM (2007)
6. Marfia, G, Rocchetti, M., Palazzi, C.E., Amoroso A.: Efficient vehicle-to-pedestrian exchange of medical data: an empirical model with preliminary results. In: Proceedings of the ACM MobiHoc Workshop on Pervasive Wireless (2011)
7. Leontiadis, I., Costa, P., Mascolo, C.: Extending access point connectivity through opportunistic routing in vehicular networks. In: Proceedings of the IEEE INFOCOM (2010)
8. Balasubramanian, A., Levine, B., Venkataramani, A.: DTN routing as a resource allocation problem. In: ACM SIGCOMM (2007)
9. Leguay, J., Friedman, T., Conan, V.: DTN routing in a mobility pattern space. In: proceedings of the ACM SIGCOMM Workshop on Delay-tolerant Networking (2005)
10. Krumm, J., Horvitz, E.: Predestination: where do you want to go today? *Computer* **40**, 105–107 (April 2007)
11. Lee, K., Yi, Y., Jeong, J., Won, H., Rhee, I., Chong S.: Max-contribution: on optimal resource allocation in delay tolerant networks. In: IEEE INFOCOM (2010)
12. Chen, Z., Kung, H., Vlah, D.: Ad hoc relay wireless networks over moving vehicles on highways. In: ACM Mobihoc (2001) p. 250
13. Bychkovsky, V., Hull, B., Miu, A., Balakrishnan, H., Madden, S.: A measurement study of vehicular internet access using in situ Wi-Fi networks. In: Proceedings of the ACM MOBICOM (2006)

14. Burgess, J., Gallagher, B., Jensen, D., Levine, B.N.: Maxprop: routing for vehicle-based disruption-tolerant networks. In: Proceedings of the IEEE INFOCOM (2006)
15. Jain, S., Fall, K., Patra, R.: Routing in a delay tolerant network. In: Proceedings of the ACM SIGCOMM (2004)
16. Yi, Y., Proutière, A., Chiang, M.: Complexity in wireless scheduling: impact and tradeoffs. In: Proceedings of the ACM MobiHoc (2008)
17. Ramanathan, R., Hansen, R., Basu, P., Rosales-Hain, R., Krishnan, R.: Prioritized epidemic routing for opportunistic networks. In: ACM MobiSys workshop on Mobile Opportunistic Networks (MobiOpp) (2007), p. 66