GeoMobCon: A Mobility-Contact-Aware Geocast Scheme for Urban VANETs

Lei Zhang, *Member, IEEE*, Boyang Yu, *Student Member, IEEE*, and Jianping Pan, *Senior Member, IEEE*

***Abstract*—Geocast, which consists of delivering messages to a specific location, has become an important technique with the accelerated development of the location-based services in mobile networks. Geocast in the automotive domain is of particular interest, enabling many promising applications, such as geographic advertising, location-based traffic alerts, etc. Different from the conventional geocast algorithms focusing on the distance-based approaches, in this paper, we propose a node mobility-aware and node contact-aware geocast algorithm (GeoMobCon) for urban vehicular ad hoc networks from the delay-tolerant network perspective to better deal with the high mobility and transient connectivity issues. Different levels and aspects of vehicle mobility information and self-maintained contact information are emp loyed, making GeoMobCon simple, scalable, and communication and computation effective. Practical issues are well considered by introducing real-world trace analysis, trace-driven simulation, and efficient buffer management. Extensive performance comparisons with other protocols have been conducted to show the advantages of GeoMobCon.**

***Index Terms*—Delay-tolerant networks (DTNs), geocast, mobility, vehicular ad hoc networks (VANETs).**

I. Introduction

In mobile networks, geocast forwards messages to destinations with specific geographic locations. The existing works of geocast protocols mainly focus on geometric distance-based approaches [1]–[4], where the distance to the destination is taken as the most important criterion for routing. However, they are not suitable for the large-scale urban vehicular ad hoc network (VANET) for the following reasons: First, because of the high node mobility and complex road structure, the distance relations of nodes change frequently and quickly, reducing the performance of the distance-based schemes; second, the schemes that require network topology information, such as GeoTORA [1] and GeoGrid [2], need to frequently update their knowledge of the network, which can cause tremendous overhead in a large-scale network, such as the urban VANET with thousands of nodes.

Manuscript received January 17, 2015; revised June 28, 2015; accepted August 2, 2015. Date of publication September 3, 2015; date of current version August 11, 2016. This paper was presented in part at the 33rd Annual IEEE International Conference on Computer Communications, Toronto, ON, Canada, April 27–May 2, 2014. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, by the Canada Foundation for Innovation, and by the British Columbia Knowledge Development Fund. The review of this paper was coordinated by Prof. A. Jamalipour.

The authors are with the Department of Computer Science, University of Victoria, Victoria, BC V8P 5C2, Canada (e-mail: leiz@uvic.ca; boyangyu@uvic.ca; pan@uvic.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2015.2476504

On the other hand, routing in delay-tolerant networks (DTNs) that specialize on intermittent connectivity is extensively studied [5], [6], [10]. We find that, even if not designed for geocast, many existing DTN routing schemes can adapt to geocast with minor modifications. However, these DTN routing schemes are originally designed for relatively small-scale networks, e.g., with up to hundreds of nodes. Concerning the scale, existing schemes either fail to achieve an acceptable performance due to the flooding-like mechanisms [5] or introduce enormous communication and computation overhead, such as the maintenance of the pairwise node contact history information [6], [10].

Instead of the traditional geometry-based approaches, we extend our previous work [8], [9] and propose a mobility-contact-aware geocast scheme (GeoMobCon) for metropolitan-scale urban VANETs from the DTN perspective, through vehicle-to-vehicle communications. Different from some of the most efficient DTN routing schemes [6], [10], which are based on the expensive pairwise contact probability calculation and sharing, our scheme employs the node mobility information at different levels, i.e., macroscopic and microscopic mobility, in addition to a relatively simple use of the contact history information. The macroscopic mobility describes the traffic trend of all vehicles in a city, whereas the microscopic mobility captures the mobility patterns of individuals. Because the macroscopic mobility for a city is relatively stable and the microscopic mobility is completely self-maintained by each vehicle, this mobility hierarchy makes our scheme distributed, simple, scalable, and communication and computation efficient when compared with existing solutions.

The two levels of mobility are extracted from the real-world GPS traces of taxicabs and buses in Shanghai, China. To facilitate mobility modeling, we divide the city into regions, each of which contains considerable traffic volumes. Traffic flows among regions are extracted and utilized as the macroscopic mobility pattern. The volume of the traffic flows can indicate how well the regions are "connected" through vehicles and how reliable the message dissemination between regions can be via vehicular communications. The massive data trace also allows us to investigate each individual's mobility pattern, which serves as the routing criterion. The proposed scheme also employs the contact history information of vehicles within the regions of interest. Considering practical restrictions, i.e., the limited buffer size and transmission bandwidth, an efficient buffer management is introduced.

The rest of this paper is organized as follows. In Section II, conventional geocast and related DTN routing schemes are discussed as related work. The trace analysis is presented in

Section III. Section IV reveals details about the proposed geocast scheme. With a trace-driven simulation, the performance is evaluated in Section V. Finally, we conclude this paper in Section VI.

II. Related Work

A. Traditional Geocast Schemes

The study of geocast has a relatively long history (since 1987) [11]. In [1] and [11], only unicast is considered. To further improve the delivery ratio, multicast, e.g., directed flooding, is widely adopted by many schemes [3], [12], [15]–[17]. Ko and Vaidya proposed two schemes [3] based on directed flooding, where one defines a rectangular forwarding zone between the source and the destination and the other forwards messages to all the neighbors who have shorter distances to the destination. While forwarding with directions, the selections of forwarding zone and nodes are important. Stojmenovic *et al.* introduced a Voronoi diagram for the forwarding zone selection [12]. It selects the nodes in a Voronoi region toward the destination as the next hop. Raw and Das proposed the work in [15], which vertically divides the vehicle transmission coverage region into two half-circular sections and selects all the border nodes as the next hop from the half-section toward the destination. Proposed by Tsiachris *et al.* and Li *et al.*, respectively, [16] and [17] select nodes near the road junction within the transmission range as next-hop vehicles.

Another category of geocast schemes uses group-based approaches [2], [4], [13], [14], where nodes are divided into groups. In GeoGrid [2], the network is partitioned into logic grids, and each partition selects one single gateway node as a group representative to forward messages. Imieliński and Navas introduced additional infrastructures as the gateways, collecting data from mobile nodes and forwarding them to the destination [4]. The maintenance of gateways introduces extra overhead.

Specifically for VANETs, different geocast schemes are proposed utilizing different information about the vehicles and the traffic to improve the performance. Li *et al.* and Yang *et al.* utilized the vehicle density and traffic load information to select the propagation route [17], [18]. Traffic lights information is used in [19]. One-hop link quality and degree of vehicle connectivity are considered in [20]. In [21], multiple metrics (i.e., distance to the destination, vehicle density, vehicle trajectory, and communication bandwidth) are considered in the routing protocol design. In [22], vehicle movement prediction in a grid road structure is used for geographic routing. However, the acquisition of such information itself could be very challenging. Leontiadis and Mascolo, as well as Cheng *et al.*, deployed navigation system for geocast purposes [13], [14]. Some applications of the geocast in VANETs are discussed in [23]–[25].

B. DTN Routing Schemes

DTNs enable communications where the source-to-destination connectivity cannot be always sustained. Some VANET systems can be typical DTNs, particularly those for nonsafety applications. Compared with the conventional geocast algorithms, DTN routing is more capable of dealing with high node mobility and transient node connectivity. For such reasons, we propose a geocast solution from the DTN's point of view in this paper.

In addition to flooding-based schemes [5], [26], another very important DTN routing category is the contact information-based routing, where a smarter relay node selection is made. In PRoPHET [7], each node maintains the encounter history with other nodes, and the routing decision is made based on the encounter probability. MaxProp [10] also utilizes the encounter information to estimate the cost of a virtual end-to-end path to the destination and uses it as the metric for routing decisions. MaxProp also takes into account realistic issues such as buffer size and bandwidth limitation. However, in [10], MaxProp is only tested on bus traces. Zhu *et al.* further took into account the intercontact time [27], [28], which are much more complicated.

III. Trace Analysis

The proposed work is based on real vehicle GPS traces collected in a modern city, i.e., Shanghai, China. There are two main benefits of introducing the real trace. First, analyzing real-world traces enables us to better understand the overall geographic distribution of the city traffic (macroscopic mobility) and individual vehicle mobility patterns (microscopic mobility). Introducing both levels of mobility along with the contact history information into the geocast scheme is the most significant novelty and contribution of our work. Second, the real-world traces enhance the reliability of our scheme by providing the realistic user mobility and contact history information.

A. About the Traces

The traces we used (partially available at http://www.cse.ust.hk/scrg) were collected from vehicles in Shanghai, i.e., 2299 taxicabs from January 31, 2007 to February 27, 2007 and 2500 buses of 103 routes from February 24, 2007 to March 27, 2007. Each bus reported a GPS report every 1 min, whereas taxicabs reported every 15 s if there was no customer on board and every 1 min when with customers. The information contained in the trace includes the vehicle ID, the latitude and longitude coordinates, timestamp, vehicle moving speed, and heading direction. In addition, taxicabs also reported whether they are hired by customers.

Different from other work taking account of only one type of vehicles in the network, in this paper, we adopt the collaboration of the different types of vehicles, i.e., taxicabs and buses. This increases the diversity of node mobility patterns, which can benefit the message dissemination for geocast. The mobility patterns considered are summarized as follows.

a) Buses: Each bus has a limited spatial and temporal coverage, i.e., moving along fixed routes during a certain period of time, which implies a very strong mobility pattern. In addition, this can be very helpful for geocast since the node mobility is somehow predictable. Given a message destination area, the buses whose routes cross that area should be preferred as message forwarders.

Fig. 1. Heat map of taxicab traffic.

b) Taxicabs: Compared with buses, the pattern of taxicab mobility is more diverse with a much larger spatial and temporal coverage. By intuition, the taxicab mobility is impacted by two factors: 1) customer demands and 2) taxicab drivers' driving habits. If one taxicab is occupied by customers, the mobility is mainly determined by the customer destination. The driver may pick the shortest path or a path with least congestion. If the taxicab is not occupied, the mobility depends on the taxi driver's driving habits and "hunting" preference.

The traces cover a large portion of Shanghai city, with the horizontal span of 70 km from the westmost record point (Qingpu District) to the eastmost point (Pudong Airport) and the vertical span of 45 km from the northmost point (Baoshan District) to the southmost one (south of Minhang District). Such a square area has a size of around 3150 km^2. However, it is not all spread with vehicles because of the city and road structures. Public vehicles appear more often in hot social spots, such as transportation hubs, commercial areas, and those regions connecting the hot spots. By counting the number of the GPS records of all taxis, we plot Fig. 1, which shows the traffic condition of the city. High traffic volume is indicated in red, whereas lower volume is indicated in green.

B. Trace Preprocessing and Region Clustering

Due to the large scale of the map, it is hard to outline all details of the regions that have distinct traffic volumes. Thus, we first discretize the map as a tiling of square cells with a size of 1 km $*$ 1 km. We let each GPS location of a report be represented by the cell to which it belongs. Then, each vehicle trajectory trace is converted to a sequence of cells. We focus on the cells with a considerable amount of traffic by counting its GPS report frequency. We are also interested in the cells, where taxi drivers pick up or drop off customers.

With the focused cells identified, we cluster them into regions. In addition, it helps to describe vehicle mobility in a proper granularity concisely. We apply the k-means clustering algorithm to these cells. The value of k determines how many regions can be formed. Since the total number of cells is fixed, with a larger value of k, more regions will be formed but with a smaller size each. Notice that clustering is only one method to define the geographic regions. Depending on different scenarios, the size, shape, and location of these regions of interest can be customized. For generality, we adopt the clustering algorithm to generate regions with similar sizes and set 40 as our default number of regions.

A traditional clustering algorithm is based on the Euclidean point distance, but we use the travel distance of two locations instead. This is because we take the real-world road structure into consideration to reflect the actual reachability between any two locations. Travel distances can be obtained through online map services, e.g., Google maps.

C. Two-Level Mobility Mining

For mobile ad hoc networks, node mobility plays a significant role in opportunistic forwarding-based routing protocols. Particularly for geographic routings such as geocast, a proper understanding and utilization of the node mobility can help to greatly improve the performance. Two levels of mobility patterns, namely, macroscopic and microscopic, are extracted from the real-world vehicle traces.

1) Macroscopic Mobility Pattern: Macroscopic mobility pattern reflects the overall traffic condition. It can be either provided by the urban planning department, or the transportation department, or obtained by counting the number of vehicles commuting between regions, etc. More of such vehicle commutes imply the stronger traffic flow between the regions, which further implies a higher reliability to transfer messages from one region to the other via vehicles. The macroscopic mobility characterizes the traffic flows between any pair of neighbor regions in the city. It reflects how regions are connected by vehicle traffic and how strong each connection is. We express the macroscopic mobility pattern by a weighted directed graph $MacMP(\mathbf{V}, \mathbf{E}, \mathbf{w})$, as shown in Fig. 3, where vertices $\mathbf{V}$ represent the regions shown in Fig. 2, the directed edges $\mathbf{E}$ represent the traffic flows, and the thickness $\mathbf{w}$ of the directed edges represents the amount of vehicle traffic, i.e., the strength of the connection. Note that the same edge connecting two regions but with different directions can have different traffic conditions, meaning with different weights for different directions. We can observe the stronger region connections in downtown areas, i.e., regions 2, 17, 23, and 32, and weaker connections on the periphery of the city, i.e., airports (regions 16 and 37). Another property of the macroscopic mobility is that such patterns are relatively stable for the whole city during different time periods of a day and different days of a month according to our observation. This is an attractive feature as such information does not need a frequent update, which greatly reduces the complexity of the scheme.

2) Microscopic Mobility Pattern: Microscopic mobility pattern, on the other hand, captures the motion patterns of individual vehicles. For buses, the ones that belong to different routes have different mobility coverage. For taxis, individuals have different mobility patterns caused by different driving behaviors of the drivers, e.g., some drivers prefer to work in downtown areas, whereas others prefer to take longer distance businesses, such as to airports. The mobility pattern can also be shown using a weighted graph similar to Fig. 3. Fig. 4(a) and (b)

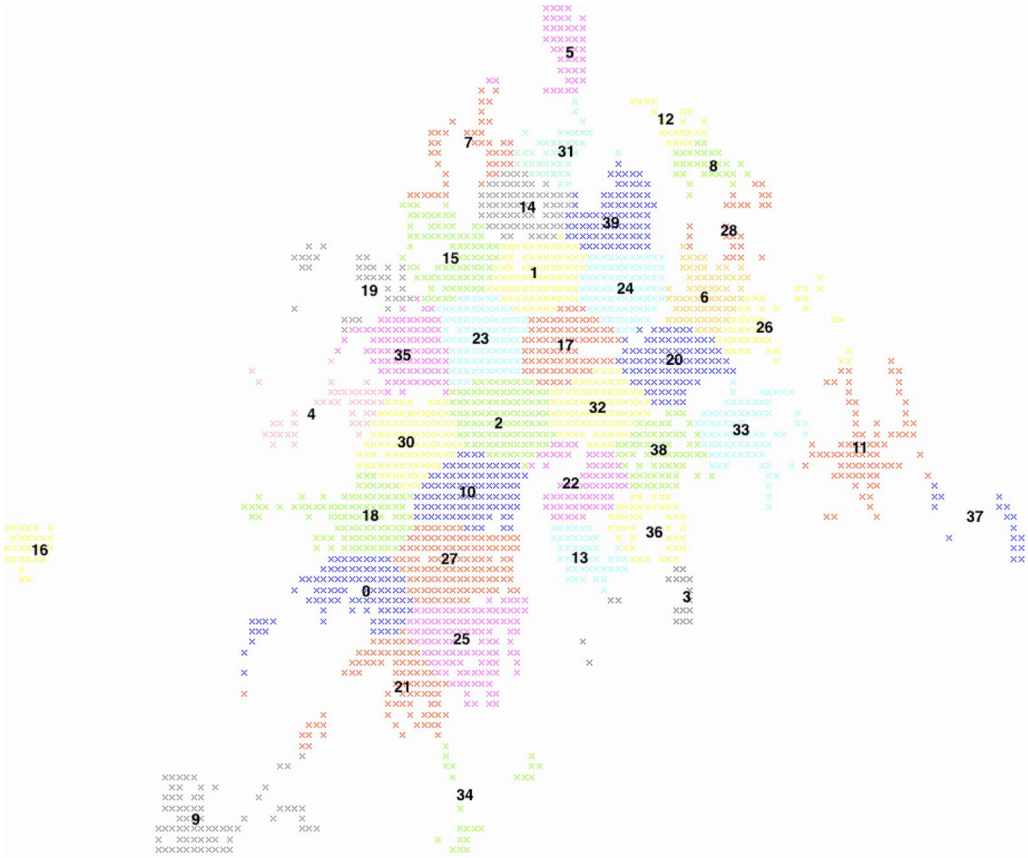


Fig. 2. Clustered regions.

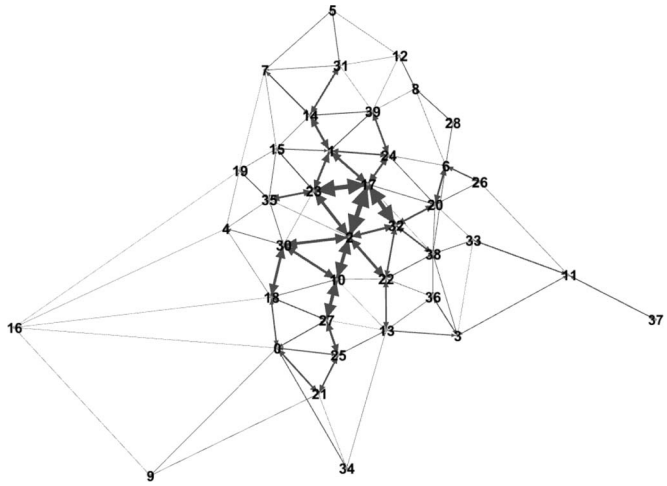


Fig. 3. Macroscopic mobility patterns.

shows the patterns of two taxis over the data collection period. Note that only edges with a considerable amount of traffic transitions are plotted for a better graph presentation. An obvious difference could be observed, i.e., Taxi 0094 was active in only downtown areas, whereas Taxi 01292 showed more activity in more regions. For a specific vehicle v, the microscopic mobility pattern can be presented as a set of conditional probabilities, i.e.,

$$MicMP_v = \bigcup P(f_i|h_n h_{n-1}, \ldots, h_1 h_0),\ f_i, h_j \in \mathbf{V},\ n \leq N$$

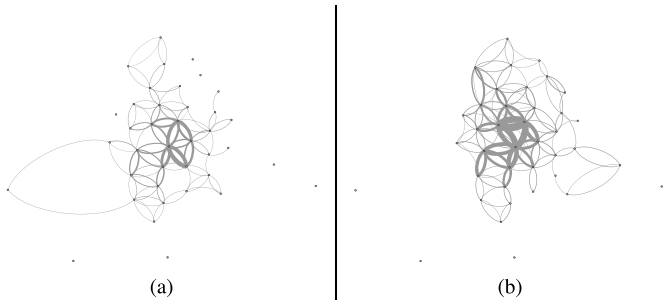


Fig. 4. Microscopic patterns for individual taxis. (a) Taxi 0094. (b) Taxi 01292.

which records all the transition probabilities to certain regions given the history information as a sequence of regions. Here, f_i indicates the region that is possible for v to go to, given that it has come from regions $h_n, h_{n-1}, \ldots, h_1, h_0$, and h_0 indicates the current region of the vehicle. If the history contains the n previous regions, we call it the nth-order conditional probability. Given a threshold N, we represent each vehicle's microscopic mobility pattern with all its possible nth-order conditional probabilities, where $n \leq N$. In addition, its space complexity is on the order of $|\mathbf{V}| * N_B^{n-1}$, where $|\mathbf{V}|$ represents the total number of regions, and N_B is the average number of neighbors for each region. The values of $|\mathbf{V}|$ and N_B can be pretty small (e.g., 40 and 4 in the given graph, respectively), leading to low space complexity.

A very desirable feature is that the microscopic mobility pattern is totally self-maintained by a vehicle itself because it only

depends on its own movement. No information sharing between vehicles is needed. Each vehicle only needs to perform a statistical analysis of its history trajectory to obtain its own microscopic mobility pattern (i.e., the conditional probabilities).

D. Mobility Entropy

To a large extent, our proposed routing scheme depends on the microscopic mobility patterns of individual taxis. It is important to understand how strong the microscopic mobility patterns are for different vehicles and how they change over time during a normal work day. To quantitatively show the activeness of individual mobility, we introduce the mobility entropy. An example is given as follows.

With clustered regions, the trip of a taxi during a certain time period can be represented as a sequence of geographic regions, e.g., for taxi A and B

$$T_A = r_3, r_2, r_2, r_3, r_5, r_2$$
$$T_B = r_1, r_2, r_3, r_5, r_4, r_1.$$

During this short period, for taxi A, the visiting frequencies of regions r_3, r_2, and r_5 are 2/6, 3/6, and 1/6, respectively. In addition, for taxi B, the visiting frequencies of regions r_1, r_2, r_3, r_4, and r_5 are 2/6, 1/6, 1/6, 1/6, and 1/6, respectively. We can observe that taxi A has a narrower mobility pattern as it moves in a limited number of regions, i.e., only three regions. On the other hand, the trace of B has higher randomness with five regions. Thus, introducing the similar concept of entropy in communication theory, the mobility entropy can be calculated as follows:

$$E_A = -\frac{2}{6}\log\frac{2}{6} - \frac{3}{6}\log\frac{3}{6} - \frac{1}{6}\log\frac{1}{6} = 0.439$$
$$E_B = -\frac{2}{6}\log\frac{2}{6} - \frac{1}{6}\log\frac{1}{6} * 4 = 0.678.$$

Taxi A has a more predictable pattern than B, which can be shown as $E_A < E_B$. We statistically studied the trace of all taxis over the whole data collection period. With different time granularity, we are able to obtain the activeness of a taxi during time periods with different lengths. For simplicity, we divide a day, i.e., 24 h, into equal-length periods, e.g., the two 12-h period scenario, the three 8-h period scenario, and the four 6-h period scenario. We compare the mobility entropy distributions of different periods in each scenario. We have found that the most distinct difference of different periods appears in the 6-h period scenario, as shown in Fig. 5. The following observations can be obtained: 1) For the first 6-h period, from 12 A.M. to 6 A.M., the entropies are much smaller than those of other periods, implying taxis are a lot less active during this period and the mobility pattern is very deterministic. 2) The entropy distributions of the other three periods are extremely similar, meaning the activeness of the taxis during their usual work hours, i.e., 6 A.M. to 12 A.M., is very similar. 3) From 6 A.M. to 12 A.M., the majority (70%) of taxis have an entropy falling between 3.5 and 4.5, showing that most taxis have similar activeness during that period of time, although the individual difference exists, as shown in Fig. 4. The mobility entropy study provides us more insights of the taxicab mobility pattern, i.e., how strong the patterns are, how they change over time, and how the pattern strength is distributed over all taxis. This is an important supplementary study to enhance our understanding of the vehicle mobility, which is the key of the proposed scheme. In the following section, the visiting frequency information will be utilized as the contact history information (with destination areas) in the routing design.

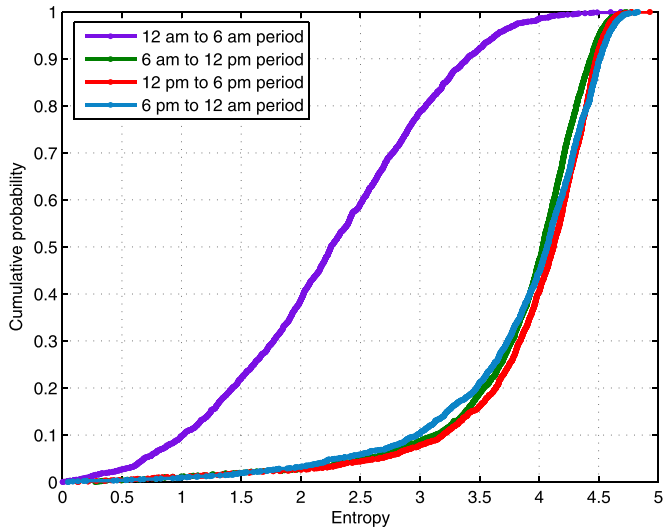


Fig. 5. Entropy distributions.

IV. Geocast Routing Scheme

Here, we propose a geocast scheme in a city scenario via taxicabs and buses. As a basic assumption in VANETs, all vehicles are cooperative for message transmission. Generated from the source node, a message is first forwarded toward the destination area through multiple hops. Once it reaches the target area, it is simply flooded within that area. Because the message flooding within the target area is relatively simple, in this paper, we mainly focus on the geographic routing from the source to the target area. This is different from the conventional unicast where the destination is a specific node. The message forwarding scheme consists of two parts: message forwarding strategy and buffer management. The node's contact history information and two levels of mobility patterns extracted are utilized in the message forwarding strategy, where the macroscopic mobility pattern is used in the forwarding path selection and the microscopic mobility patterns are used in routing decision making.

A. Message Routing Strategy

The key idea of GeoMobCon message routing is to utilize the two levels of vehicle mobility and the contact history information. An optimal (i.e., strong region-to-region connectivity via vehicle traffic) routing path, which is a sequence of regions toward the destination, is selected based on the macroscopic mobility when a message is generated. Then vehicles try to forward this message along such a path toward the destination. A vehicle's microscopic mobility and its contact history with the destination determine whether it can help to forward the message to its next hop or further hops on the optimal path. Notations used are given in Table I.

TABLE I
NOTATIONS

Notation	Explanation
$\mathbf{FR}$	A list of a vehicle's future regions
$\mathbf{TG}$	A list of a message's ToGo regions
M_i	Message with ID i
LM_m^v	Vehicle v's mobility-based forwarding likelihood of message m
LC_r^v	Vehicle v's contact-based visiting likelihood of region r
N	Threshold for the number of history regions
$h_n h_{n-1} \cdots h_1 h_0$	The nth-order history regions of a vehicle
C_{age}	Aging constant for the contact-based visiting likelihood
$C_{encounter}$	Initialization constant for the contact-based visiting likelihood

1) Macroscopic Mobility-Based Forwarding Path Selection: It is believed that the larger traffic volume toward a region implies a better connectivity to that region. The chance of successfully delivering a message to that region is also higher. Thus, we start our geocast routing by first selecting an optimal forwarding path leading to the destination with the best connectivity based on the graph shown in Fig. 3. Once a message is generated, the source calculates the optimal path leading to the destination. Such an optimal path is a region sequence that has the best connectivity (i.e., in terms of the traffic volume) from the origin to the target destination, and the path information is embedded in the message header once the path is calculated. To find such a path, Dijkstra's algorithm is applied to the weighted graph in Fig. 3, where the weight on the edge is determined by the total traffic volume between regions. All vehicles should preload the overall traffic weight graph to be capable of calculating the optimal path for each message. Since the city traffic pattern does not change dramatically over time, there is no need to frequently update this graph, which greatly reduces the overhead for vehicles. Vehicles try their best to forward the messages along their optimal paths to increase the delivery ratio. The message is transferred in a region-by-region manner, i.e., once the message enters one region, suitable vehicles are selected to forward it to the next region along the optimal path.

2) Microscopic Mobility-Based Routing Decision: Each vehicle also maintains its own mobility pattern, i.e., the microscopic mobility pattern. Individual mobility pattern is used for making explicit routing decision when two vehicles encounter each other. As mentioned in Section III, such a pattern is represented as a collection of conditional probabilities. These conditional probabilities can be obtained from the movement history and further updated as vehicles move among the regions.

According to the current vehicle location, relay vehicles will try to forward a message to the next optimal region according to the optimal path stored in the message. We use the term *likelihood* to determine how capable a vehicle is to forward a message to the next optimal region. Assume the optimal path for message m is $r_2 \rightarrow r_3 \rightarrow r_7 \rightarrow r_5 \rightarrow r_6$, where r_6 indicates the destination region. When two vehicles encounter each other in region r_5, one of them, e.g., v_i, is carrying m, and the other, e.g., v_j, is not. Then, v_i first requests v_j's mobility-based delivery likelihood LM_m^j for message m. If $LM_m^j > LM_m^i$, v_i forwards the message to v_j; otherwise, the contact history information will be used for the routing decision making. This process is expressed in Algorithm 1.

Algorithm 1 Routing Strategy

```
1:  procedure ROUTINGSTRATEGY (vehicle E, R, message m)
2:      LM_m^E = LikelihoodMobUpdate(E, m)
3:      LM_m^R = LikelihoodMobUpdate(R, m)
4:      D is the destination region of m
5:      LC_D^E = LikelihoodConUpdate(E, D)
6:      LC_D^R = LikelihoodConUpdate(R, D)
7:      if LM_m^E < LM_m^R then
8:          if LM_m^R > 0 then
9:              E forwards m to R
10:             if LM_m^E < 0 then
11:                 Drop m
12:             end if
13:         else
14:             E does not forward m to R
15:         end if
16:     else
17:         if LC_D^E < LC_D^R then
18:             E forwards m to R
19:             if LM_m^E < 0 then
20:                 Drop m
21:             end if
22:         else
23:             E does not forward m to R
24:             if LM_m^E < 0 then
25:                 Drop m
26:             end if
27:         end if
28:     end if
29: end procedure
```

Again, a vehicle maintains a likelihood for each message it is carrying, and these likelihoods are used to compare with other vehicles' likelihoods to make the routing decision. Algorithm 2 gives the details of how the likelihood is calculated and updated. In Algorithm 2, we utilize the near-future location information if it is available. This is because we consider the mobility features of city buses and taxis, whose near-future location may be easily accessible. We denote the near-future regions of a vehicle, if available, as "future regions," i.e., $\mathbf{FR} = \{fr_1, fr_2, \ldots, fr_n\}$ and the regions the message still has to go through according to its optimal path as "ToGo" regions, i.e., $\mathbf{TG} = \{tg_1, tg_2, \ldots, tg_n\}$. The forwarding decision process (see Algorithm 1) and the likelihood update logic (see Algorithm 2) are summarized as follows.

Algorithm 2 Mobility-Based Likelihood Update

```
1: procedure LIKELIHOODMOBUPDATE(vehicle V and message m)
2:     if V's FR is known then
3:         if V's FR intersects with m's TG then
4:             Find the intersection region I closest to the destination
5:             Record I's index in TG idx
```

$LM_m^V = 1 + (idx/TG.length)$
else
$LM_m^V = -1$
end if
else
Start nth-order prediction, $n = N$
while nth-order pattern does not exist in pattern set **do**
$n = n - 1$
if $n < 2$ **then**
$LM_m^V = -1$
Break
end if
end while
$LM_m^V = P(TG_1|h_n h_{n-1}, \ldots, h_1 h_0)$
end if
Return LM_m^V
end procedure

If the encountered vehicle's near-future information is known, such information is utilized. For taxis carrying customers, their destinations are determined by the customers on board. For buses, because they have prefixed routes, their future regions over a short period are predictable. For such vehicles, we only need to see if their future locations have any intersections with the message's desired future optimal regions. If yes, these vehicles are definitely helpful to forward that message. Otherwise, the message is not passed to them. There are also vehicles whose destinations are unknown, such as taxis without customers. At this point, the taxi's microscopic mobility pattern becomes the source for location prediction. According to v_j's historical mobility information, it replies to v_i with the probability of going to the optimal region given the current trajectory history, as v_j's forwarding likelihood, according to the conditional probability from its microscopic mobility pattern. If v_j has a high enough probability, i.e., higher than v_i's likelihood, to go to any of the future optimal regions of the message, v_i will forward the message to v_j. Otherwise, that specific message will not be forwarded.

3) Contact History-Based Routing Decision: It is also possible that there are some vehicles that are less likely to go to the expected future regions, except for the destination region. These vehicles can also have a high chance to deliver the message. However, these vehicles will be excluded by the mobility-only-based routing process mentioned before. To grab such forwarding opportunities, we introduce another routing decision metric, i.e., contact history, in addition to the microscopic mobility-based routing decision–making strategy. As shown in Algorithm 1, when a message carrier v_i compares the mobility-based likelihoods of itself and the encounter v_j and notices that $LM_m^j < LM_m^i$, it will request v_j's contact history-based likelihood $LC^j_{\text{destination}}$. If $LC^j_{\text{destination}} > LC^i_{\text{destination}}$, v_i still forwards the message to v_j.

Each vehicle maintains its contact history information with each region of interests. Such history information is updated in real time, meaning whenever it visits a region of interest, the contact history information with that region is updated. We adopt the contact history updating method in [6], with the consideration of the aging factor. Define LC_d^j as the contact-based visiting likelihood of vehicle v_j and region of interest r_d. The initial value of it is set to be 0. Whenever v_j visits r_d, it updates the value of LC_d^j, i.e.,

$$LC^j_{d\,(\text{new})} = LC^j_{d\,(\text{old})} + \left(1 - LC^j_{d\,(\text{old})}\right) * C_{\text{encounter}}$$

where $C_{\text{encounter}} \in (0, 1)$ is a constant, and according to [7], we set it to be 0.75. We also consider the aging factor, which means the value of LC_d^j is reduced with time elapsing since the last visit. In addition, this can be represented in the following equation:

$$LC^j_{d\,(\text{new})} = LC^j_{d\,(\text{old})} * C^k_{\text{age}},$$

where $C_{\text{age}} \in (0, 1)$ is the aging constant and set to be 0.98 [7]. k represents the number of time units elapsed. This updating process is summarized in Algorithm 3. Note that the contact history information is updated in real time, which means whenever a vehicle visits a region, it updates the contact history information to that region. Since the number of the regions of interest is limited, such information maintenance is scalable with space complexity on the order of $|\mathbf{V}|$, where $\mathbf{V}$ is the set of the regions of interest, as defined in Section III.

Algorithm 3 Contact-Based Likelihood Update

procedure LIKELIHOODCONUPDATE(vehicle V and region D)
k is the time units elapsed since the last time V visited D
$LC^V_{D\,(\text{new})} = LC^V_{D\,(\text{old})} * C^k_{\text{age}}$
if V enters D **then**
$LC^V_{D\,(\text{new})} = LC^V_{D\,(\text{old})} + \left(1 - LC^V_D\right) * C_{\text{encounter}}$
end if
Return $LC^V_{D\,(\text{new})}$
end procedure

B. Buffer Management

To be realistic, we assume each vehicle carries a limited-size buffer on board and has limited wireless transmission bandwidth. Thus, when two vehicles encounter, not all the messages in the buffer can be transmitted at one contact. To effectively utilize the buffer space and each contact opportunity, the buffer management of our routing scheme involves the following mechanisms.

First, after a message carrier V forwards the message m to another vehicle, it will check its likelihood LM_m^V. For each message m and vehicle V pair, LM_m^V changes with the location of the vehicle. If LM_m^V becomes too small, e.g., below zero, and it has already been forwarded to others, the vehicle deletes the message to reduce the extra storage and communication overhead.

Second, each vehicle prioritizes the messages in the transmission queue. When vehicle v_i encounters v_j, v_i may have more than one message suitable to be forwarded to v_j. Before the transmission starts, v_i first sorts the messages in its transmission

TABLE II
COMPARISON AMONG MULTIPLE SCHEMES

Routing algorithm	Abbreviation	Knowledge for routing decision	Buffer management
GeoEpidemic [26]	GE	None	No
GeoPRoPHET [7]	GP	Contact history	No
GeoMaxProp [10]	GMx	Contact history	Yes
GeoDist	GD	Distance to the destination	Yes
GeoMob [8]	GMb	Node mobility	Yes
GeoMobCon	GMC	Node mobility & contact history	Yes
GeoMobCon-NoBufMgt	GMCNoBM	Node mobility & contact history	No

queue according to the likelihood of L_m^j descendingly. Note that the messages are sorted according to the message delivery likelihood from v_j's point of view so that the message that can take the most benefit from v_j gets transmitted to v_j first.

Third, the acknowledgments of the received message are flooded in the network to remove the redundant copies of the message. We borrowed a similar idea from MaxProp [10], where each acknowledgment is a 128-bit hash of the received message ID, source, and destination. The cost is little if the acknowledgment is small compared with data packets, which is 512 kb in our simulation setting. In addition to the small buffer size overhead, it also costs little communication overhead, which has been shown in [10] that no more than 1% of the average connection duration is spent on sending acknowledgments.

V. PERFORMANCE EVALUATION

A. Protocol Comparisons

We compare our scheme with three other DTN-based geocast protocols and one traditional distance-based protocol. Although the DTN-based ones are not originally designed for geocast, by setting the destination with stationary mobility, we can easily convert them for the geocast purpose. To distinguish from the original schemes, we add a "Geo" prefix to the original scheme names. Short descriptions about the schemes are given as follows.

- **GeoEpidemic** [26]: Whenever two vehicles encounter each other, they exchange as many messages as they can, which is subject to buffer and bandwidth constraints. This is a simple flooding.
- **GeoPRoPHET** [7]: PRoPHET makes routing decisions based on the history of contact information. Messages are forwarded to the nodes who have a higher contact frequency with the destination. A *transitive* property is considered when updating the contact frequency. If A meets B frequently while B meets C frequently, then it is believed that A and C should have a relatively high contact frequency.
- **GeoMaxProp** [10]: Similar to PRoPHET, MaxProp also utilizes the history of contact information. Each node maintains a contact graph based on the history of contact information of both its own and its encounters'. Routing decisions are made according to the cost of a delivery path going through a specific neighbor. A low-cost delivery acknowledgment scheme is adopted, and the transmission queue is prioritized considering both message hop counts and delivery probabilities.
- **GeoDist**: We compare our scheme with a distance-based geocast scheme (i.e., GeoDist), which is the key algorithm for many existing geocast schemes. Messages are forwarded to the nodes with a shorter distance to the destination area. Similar to [29], for possible transmissions, the messages in the transmission queue are prioritized according to the "improvement" that can be possibly achieved by a transmission. The "improvement" here means the reduction of the distance to the destination. The greater the "improvement," the higher the priority assigned, and the sooner the message can be transmitted.
- **GeoMob** [8]: GeoMob [8] is the precursor of the currently proposed GeoMobCon. In GeoMob, the same buffer management scheme as GeoMobCon is adopt. However, for the routing decision, only the two-level mobility model is utilized. Contact history information between vehicles and regions is not utilized. This is the main difference from its enhanced version, i.e., GeoMobCon.
- **GeoMobCon-NoBufMgt**: Finally, to show the effectiveness of the buffer management in our proposed scheme, we add a modified version of GeoMobCon, i.e., GeoMobCon-NoBufMgt, which removes the buffer management mechanism, for comparison. It has the mobility-based and contact history information-based routing strategy but no buffer management.

A scheme comparison is shown in Table II. Flooding-based schemes, e.g., Epidemic, do make no routing decisions to select relay nodes. The other schemes carefully select relay nodes based on different information, i.e., contact history or distance information, leading to a controllable amount of copies for each message. The amount of copies varies in different situations. For instance, using GeoPRoPHET, if the destination is active and has rich contact history with many nodes, a large number of copies are expected to be forwarded to those nodes. However, if the destination is not very popular, a limited number of copies will be generated. Among all schemes, our scheme, as well as GeoMaxProp, GeoDist, and GeoMob, has taken into account the practical restriction of the buffer size and transmission bandwidth, which is reflected by introducing the buffer management mechanisms.

We provide a further comparison between other schemes and ours. First, traditional distance-based geocast routing, e.g., GeoDist, is limited to a large-scale urban vehicular network. With the complex road structure and high vehicle mobility,

the relative position of vehicles changes quickly. A vehicle, which is currently closer to the destination, may be farther away at the next second, because either it takes a detour toward another area, it is moving in the opposite direction, or it just stops. Thus, taking the distance as the only routing criterion is not sufficient and suitable for the city vehicle network environment.

Second, as compared with routing schemes PRoPHET and MaxProp, our scheme is more distributed and requires much less information from other nodes. As described, in PRoPHET, the routing decision for one node is solely based on its contact history with the destination. To maintain its own encounter probability with the destination, each node has to frequently acquire its neighbors' encounter probabilities, whenever there is a chance. Similarly, MaxProp requires sharing every node's contact history information among all nodes to maintain the contact graph up to date. Remember that the contact history information is always changing as nodes move and meet each other; thus, an enormous communication overhead will be introduced, not to mention if the network scale is large, such as with thousands of nodes in the city case. To manage and update all the contact history information in the local buffer, e.g., frequently searching, updating, etc., also introduces computation overhead. Most existing work did not count it as the scheme overhead since it is on the control plane. However, the complexity issue can become severe in the real world.

Our scheme also requires global information, i.e., the macroscopic mobility pattern; however, the city macroscopic mobility pattern stays quite stable over time according to our observation of the trace. Unless there are some major changes happening, e.g., a new highway is built, which changes the city traffic dramatically, there is no need to frequently update the macroscopic mobility pattern. The other information needed in our scheme is the microscopic mobility and local contact history information, which are totally self-maintained by each node. This is a very attractive feature for a large-scale distributed network. GeoMob and GeoMobCon-NoBufMgt are also listed. Comparing with GeoMobCon, they do not have the contact-based routing strategy and buffer management, respectively.

B. Simulation Setup

We use a Java-based open-source simulator ONE [30] for simulation. To be practical, we randomly select 200 taxis and 300 buses, importing their traces for the node mobility.

According to the specifications of IEEE 802.11p [31], [32], we set the vehicle transmission speed to 6 Mb/s. The buffer size is set to 2000 MB. We test our protocol with different simulation settings to give a comprehensive understanding of the impact of different factors. For different transmission power and urban environment, we test our protocol performance under different transmission ranges, i.e., ranging among 50, 250, 500, and 1000 m. Different applications may have different message generation rates and time-to-live (TTL) periods, and thus, we test our protocols with different message generation intervals, varying among 10, 60, and 300 s, and different message TTLs, varying among 60, 120, and 180 min.

Messages are randomly generated among all taxis, and the destination region is randomly chosen. We set the simulation time to 24 h with the first 6 h as a warm-up period.

C. Message Delivery Performance

With different parameter combinations, the performance changes. To understand the overall performance, we study two extreme cases. *Pessimistic case*: All parameters are set to restrain the message transmission, e.g., the transmission range is set to the smallest 50 m, the message generation interval is the lowest 10 s, and the TTL is set to be the lowest 60 min. *Optimistic case*: All parameters are set to smooth the message transmission, e.g., the transmission rage is set to be the largest 1000 m, the message generation interval is the largest 300 s, and the TTL is set to be the highest 180 min. We are interested in how the system performs under these two extreme conditions.

For comparison, four performance metrics are discussed, including the following: delivery ratio, which is calculated as N_D/N_G, where N_D is the total number of delivered messages, and N_G is the total number of generated messages; overhead ratio, which is calculated as $(N_R - N_D)/N_D$, where N_R is the total number of message relays; average latency, which is the average delay for successful deliveries; and average hop count, which is the average hop count for the delivered messages. Because the performance of the geocast is dominated by the geographic forwarding process (i.e., from the source to the target region), the aforementioned performance metrics are all calculated for this phase. Once a message reaches the target region, we assume a simple flooding scheme will be applied to all schemes within the destination region. Thus, the flooding performance should be the same for all schemes.

1) Impact of the Transmission Range: Table III lists the comparison results with the change of the transmission range, with the top section of the table illustrating the results for the pessimistic case and the bottom section for the optimistic case. As we can see, the delivery ratio, overhead ratio, and average hop count for all protocols in both cases increase with the rise of the transmission range. This is because, the larger the transmission range, the further the message can be forwarded to within one hop. Thus, the message can reach the destination more easily. This leads to the increase of the delivery ratio. With a larger range, more relay candidates show up for each transmission, leading to an increase of the overhead. With the increase of the delivery ratio, more messages are delivered to the destination than the scenario with a smaller transmission range. According to the simulations, when the transmission range is small, the delivery ratio is low, and most of the delivered messages are generated at locations close to their destinations. Therefore, the number of hops required is usually low. However, with a larger transmission range, more messages get transmitted successfully to their destinations, which are much further away from their origins. In particular, when the distance between the destination and the origin is much larger than the transmission range, which is our case because of the large city considered, more hops are needed for successful transmissions, leading to the increase of the average hop count.

In terms of the average latency, for the pessimistic case, we see for all protocols that the latency first increases then decreases. When the transmission range is small, vehicles have much fewer contacts with other vehicles, therefore, a large portion of the delivered messages are those whose destinations are close to their sources. In particular when the TTL is short, which is the fact in the pessimistic case, carry-and-forward is the main approach to deliver these messages. Thus, these messages are easily delivered to their nearby destinations with a relatively short latency. However, the delivery ratio in such a case is very low. However, with the increase of the transmission range, more messages can be delivered, but with a longer travel distance and a longer latency. Thus, the average latency increases. When the transmission range continues increasing, more messages can be delivered with multihop transmissions, which is much faster than the carry-and-forward transmission, making the average latency drop. On the other hand, for the optimistic case, the average latency decreases with the increase of the transmission range. This is because with a larger TTL (i.e., 180 min for the optimistic case), messages have longer time to be propagated to the destinations, even when the transmission range is small. Then for a larger transmission range, messages get more chances to be delivered, resulting in a shorter latency.

We can also observe that, in terms of the delivery ratio and average latency, for different transmission ranges, our proposed protocol GeoMobCon performs slightly worse than GeoEpidemic and GeoMaxProp. However, we have to emphasize that, unlike GeoMaxProp utilizing global control information, which implies a large control overhead, our protocol only utilizes the local information (mobility and contact history information) of each node, making it more scalable and practical. In addition, different from GeoEpidemic and GeoDist with a flooding fashion, our protocol makes smarter relay selection, so that it achieves a very low overhead ratio (up to 40% less than GeoEpidemic, GeoPRoPHET, GeoMaxProp, and GeoDist, on average). Our proposed protocol also achieves a low average hop count. This implies that our protocol requires fewer transmissions, leading to less power consumption and causing less interference. The proposed GeoMobCon also outperforms its precursor GeoMob, particularly in delivery ratio (up to 30% higher).

2) Impact of the Network Traffic Intensity: Table IV shows the impact of the network traffic intensity, which is represented by the message generation interval, from 10 s per message, 60 s per message, to 300 s per message. With the increase of the message generation interval (i.e., decrease of the message generation rate), we see dramatic changes in the pessimistic cases. With less traffic intensity, there is less competition for message transmissions, leading to the increase of the delivery ratio. However, when the parameter setting is harsh for the message transmission, contact opportunity becomes precious, and the mobility-based routing decision-making strategy (e.g., GeoMob, GeoMobCon, and GeoMobCon-NoBufMgt) can suppress the efficiency of using the contact opportunities, leading to the fluctuation of the delivery ratio.

From the results of the overhead ratio for both pessimistic and optimistic cases, we see the overhead ratio increases with the increase of the message generation interval. This is because the network overhead is defined as $(N_R - N_D)/N_D = (N_R/N_D) - 1$. When the message generation interval increases, the network traffic is reduced. Therefore, with a lower traffic competition, each message gets more chances to be transmitted to more vehicles, which increases the total number of relays. The trend for average latency results differs for the pessimistic and optimistic cases. For the pessimistic case, because of the harsh setting for the message transmission, when the traffic intensity is high, most of the delivered messages are those that are easy to deliver, such as those whose source locations are close to the destinations. When the traffic intensity is low, each message has a higher chance of being forwarded through more hops (i.e., the increase of the average hop count and the overhead ratio), leading to the increase of the average latency. However, in the optimistic case, because of the larger transmission range and longer TTL, the majority of the messages get delivered. Therefore, the delivery ratio, overhead ratio, average latency, and average hop count all get increased compared with the results of the pessimistic case. In particular, for the latency, lower traffic intensity reduces the transmission competition, leading to the decrease of the average latency.

In terms of the protocol comparison, for the similar reasons mentioned before, i.e., being local-information-based and more scalable, GeoMobCon performs slightly worse in terms of the delivery ratio and average latency than GeoEpidemic and GeoMaxProp but much better than GeoMob. In addition, its simplicity and scalability make it achieve the lower overhead ratio (decreased by 39%) and average hop count (decreased by one hop) compared with the average of all GeoEpidemic and GeoMaxProp scenarios.

3) Impact of Message TTL: The impact of TTL is demonstrated in Table V. With a longer TTL, messages have longer time to stay in the network and be transmitted, leading to a higher delivery ratio, larger average latency, and average hop count. With more messages delivered, the delivery efficiency becomes higher, leading to the drop of the overhead ratio. Comparing the two cases, with less congested traffic and a larger transmission range, the results of the optimistic case show a higher delivery ratio, overhead ratio, and average hop count. However, the average latency is reduced since the optimistic case has more contact opportunities for the message propagation. To compare the performance of different protocols, similar conclusions as previous discussions on the other impacting factors can be drawn.

4) Evaluation of Proposed Design Components: As mentioned, GeoMob, GeoMobCon, and GeoMobCon-NoBufMgt are all mobility-based routing protocols. Comparing with other protocols, the mobility-based protocols are featured with the low overhead and highly distributed manner. GeoMob is the original protocol proposed in our previous work [8], including two major components: mobility-based routing strategy and buffer management. On top of it, we propose its extended version GeoMobCon, considering the contact history for routing. As we can see from Tables III–V, in most scenarios, GeoMobCon achieves the same low overhead ratio as GeoMob, which is the lowest of all. However, at the same time, it outperforms GeoMob in terms of the delivery ratio and average latency. The improvement shows the advantages of the contact

TABLE III
IMPACT OF THE TRANSMISSION RANGE

		Delivery Ratio				Overhead Ratio				Average Latency				Average Hop Count			
		50 m	250 m	500 m	1000 m	50 m	250 m	500 m	1000 m	50 m	250 m	500 m	1000 m	50 m	250 m	500 m	1000 m
Pessimistic Case	GE	0.0981	0.1964	0.2902	0.4593	80.1545	206.1503	277.1300	322.7495	1742.1816	2073.0041	2052.0399	1813.2913	2.0083	3.1662	4.0945	7.5983
	GP	0.0950	0.1672	0.2428	0.3853	33.3715	101.8007	168.4447	244.6428	1650.3045	1949.2138	1992.8980	1880.0270	1.7674	2.6671	3.4166	5.1826
	GMx	0.1110	0.2221	0.3154	0.4873	68.1689	174.8395	243.6363	365.0463	1807.0219	2053.3580	2026.2422	1734.0513	2.0667	3.2110	4.1802	8.3038
	GD	0.0965	0.1810	0.2635	0.3884	68.7818	192.2653	262.9886	329.1949	1762.3429	2050.4834	2087.5635	1925.4774	1.9197	2.9987	3.8054	5.8546
	GMb	0.0764	0.1260	0.1840	0.3020	37.9909	106.3903	166.9239	250.0272	1565.1030	1907.7153	2010.4327	1999.0241	1.5318	2.3581	2.9931	4.3473
	GMC	0.0921	0.1642	0.2303	0.3693	38.3706	101.4313	161.8030	252.7957	1682.1005	1946.3622	1978.1859	1823.3457	1.8543	2.9154	3.6196	6.1228
	GMCNoBM	0.0921	0.1637	0.2296	0.3704	41.7613	115.5806	184.8826	257.1003	1664.5126	1955.7992	1984.4264	1836.2519	1.8467	2.8168	3.5393	5.9731
Optimistic Case	GE	0.5313	0.6597	0.7292	0.8021	233.1307	281.4105	275.8048	272.8485	6077.4771	5006.0211	4417.2762	3161.4719	4.4444	5.4526	6.5857	13.1082
	GP	0.3715	0.5833	0.6701	0.7535	125.0280	217.3631	238.9948	255.3779	6267.6075	5698.4048	4854.3834	3857.7880	3.4579	4.5238	5.4819	7.6175
	GMx	0.5521	0.6736	0.7326	0.8125	174.9560	218.8299	270.2844	526.0641	5922.8931	4998.1443	4449.3460	3241.9231	4.4214	5.4021	6.6114	13.3846
	GD	0.5174	0.6493	0.7222	0.7743	208.3691	273.8289	271.8702	275.4260	6168.8859	5154.4278	4784.4615	3863.2646	4.4228	5.1872	5.6827	8.2377
	GMb	0.3333	0.5243	0.6285	0.7257	94.5417	167.0861	185.8122	252.7656	6372.1667	5951.0728	5393.3370	4590.0287	3.1875	4.1258	4.4420	5.8182
	GMC	0.3924	0.5556	0.6424	0.7465	94.6106	169.8250	198.0595	276.7767	6039.2212	5614.6750	5162.2919	4133.8791	3.6903	4.5188	4.9838	7.3767
	GMCNoBM	0.3924	0.5590	0.6458	0.7500	143.0619	240.8199	253.2151	253.1713	6024.2124	5635.1304	5096.5376	4079.0741	3.5044	4.5528	4.9355	7.3287

TABLE IV
IMPACT OF THE NETWORK TRAFFIC INTENSITY

		Delivery Ratio			Overhead Ratio			Average Latency			Average Hop Count		
		10 s	60 s	300 s	10 s	60 s	300 s	10 s	60 s	300 s	10 s	60 s	300 s
Pessimistic Case	GE	0.0981	0.1319	0.1389	80.1545	154.2421	211.1250	1742.1816	1986.8947	2315.3500	2.0083	2.7526	2.9750
	GP	0.0950	0.0924	0.0833	33.3715	54.9624	72.5417	1650.3045	1766.2857	2029.7500	1.7674	1.8722	2.3750
	GMx	0.1110	0.1417	0.1424	68.1689	135.1176	184.9024	1807.0219	1987.3235	2350.6341	2.0667	2.7402	3.0000
	GD	0.0965	0.1181	0.1354	68.7818	138.4706	176.0256	1762.3429	1910.2235	2326.6667	1.9197	2.5059	2.9744
	GMb	0.0764	0.0799	0.0694	37.9909	53.3739	64.3000	1565.1030	1753.4609	2006.5000	1.5318	1.7913	2.1000
	GMC	0.0921	0.0993	0.0868	38.3706	55.0140	71.5200	1682.1005	1846.1678	2065.6000	1.8543	2.1329	2.2400
	GMCNoBM	0.0921	0.1000	0.0868	41.7613	61.1528	79.9600	1664.5126	1837.7361	2065.6000	1.8467	2.1181	2.2400
Optimistic Case	GE	0.7606	0.7958	0.8021	272.2491	275.3805	272.8485	3893.8841	3206.3421	3161.4719	7.6363	11.9075	13.1082
	GP	0.7209	0.7465	0.7535	228.4184	255.5860	255.3779	4257.5251	3874.9916	3857.7880	5.3066	7.1330	7.6175
	GMx	0.7851	0.7979	0.8125	294.5212	490.3455	526.0641	3572.2506	3143.5126	3241.9231	7.9244	12.1784	13.3846
	GD	0.7324	0.7646	0.7743	270.5509	278.7947	275.4260	4339.5632	3823.0372	3863.2646	6.1862	7.9228	8.2377
	GMb	0.6819	0.7063	0.7257	203.4725	244.5497	252.7656	4684.7766	4459.7345	4590.0287	5.0375	5.7207	5.8182
	GMC	0.7080	0.7264	0.7465	214.9053	268.4771	276.7767	4204.6860	4005.1511	4133.8791	6.0947	7.1941	7.3767
	GMCNoBM	0.6993	0.7333	0.7500	245.6610	257.5843	253.1713	4361.3595	4039.5795	4079.0741	5.6900	7.1203	7.3287

TABLE V
IMPACT OF THE MESSAGE TTL

		Delivery Ratio			Overhead Ratio			Average Latency			Average Hop Count		
		60 min	120 min	180 min	60 min	120 min	180 min	60 min	120 min	180 min	60 min	120 min	180 min
Pessimistic Case	GE	0.0981	0.1619	0.2081	80.1545	64.6112	54.9772	1742.1816	3709.6898	5521.6885	2.0083	2.1229	2.1835
	GP	0.0950	0.1684	0.2133	33.3715	32.6323	31.7238	1650.3045	3489.1162	5101.6408	1.7674	2.0344	2.0119
	GMx	0.1110	0.1999	0.2728	68.1689	50.5553	40.5342	1807.0219	3743.0782	5532.1672	2.0667	2.2942	2.3823
	GD	0.0965	0.1575	0.2054	68.7818	58.2344	49.2648	1762.3429	3666.1470	5492.7639	1.9197	2.1212	2.1544
	GMb	0.0764	0.1374	0.1880	37.9909	39.3328	35.9470	1565.1030	3534.9722	5342.4532	1.5318	1.8332	1.9938
	GMC	0.0921	0.1718	0.2324	38.3706	35.7028	31.6414	1682.1005	3655.9838	5478.1624	1.8543	2.1833	2.3710
	GMCNoBM	0.0921	0.1681	0.2204	41.7613	39.7266	36.7553	1664.5126	3649.4091	5387.5179	1.8467	2.1081	2.1780
Optimistic Case	GE	0.5069	0.7326	0.8021	347.4589	285.6588	272.8485	1545.1644	2638.3981	3161.4719	12.6986	13.2512	13.1082
	GP	0.3993	0.6667	0.7535	297.9826	261.9375	255.3779	1884.6783	3212.0521	3857.7880	6.7826	7.5052	7.6175
	GMx	0.5069	0.7326	0.8125	505.9452	505.3981	526.0641	1550.8082	2629.8578	3241.9231	13.0068	13.5355	13.3846
	GD	0.4028	0.6840	0.7743	363.0948	291.2437	275.4260	1843.7931	3208.2335	3863.2646	8.1466	8.0964	8.2377
	GMb	0.3090	0.5938	0.7257	291.7640	258.4620	252.7656	2158.9663	3660.2924	4590.0287	4.9551	5.4854	5.8182
	GMC	0.3750	0.6250	0.7465	303.7037	283.0333	276.7767	1793.5741	3232.5000	4133.8791	7.3889	7.3333	7.3767
	GMCNoBM	0.3785	0.6389	0.7500	286.7339	267.1250	253.1713	1776.8807	3244.8261	4079.0741	7.3486	7.3370	7.3287

history-based routing strategy. Because the contact history of each vehicle is also self-maintained, it does not conflict with the highly distributed feature.

On the other hand, when comparing GeoMobCon with GeoMobCon-NoBufMgt, we can see the benefit of the buffer management. From the tables, we can observe that GeoMobCon-NoBufMgt performs very close to GeoMobCon in different scenarios. However, in the optimistic cases, GeoMobCon has less overhead (up to around 33% less). This is mainly because buffer management can effectively remove

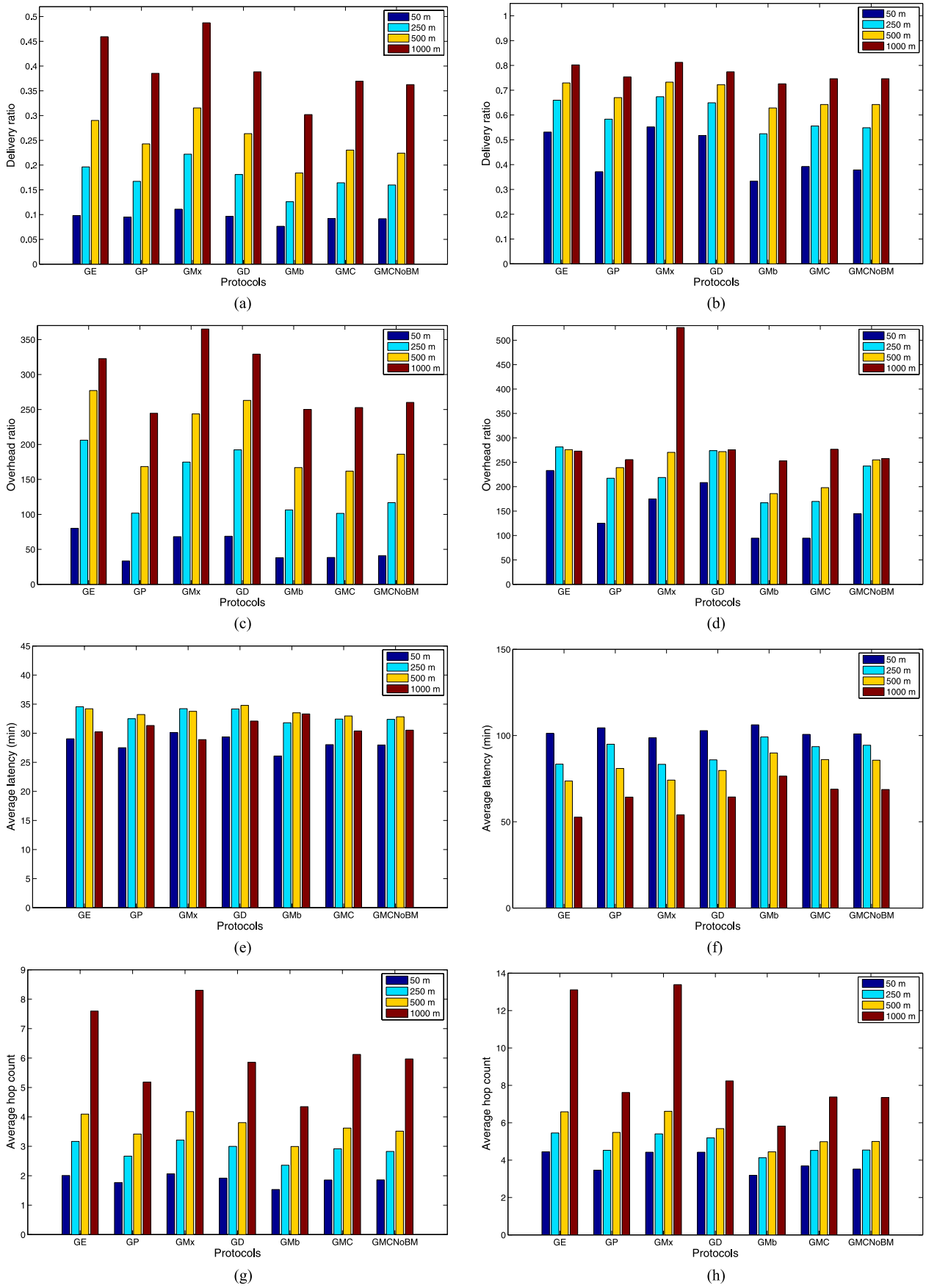


Fig. 6. Impact of the transmission range. (a) Delivery ratio, pessimistic case. (b) Delivery ratio, optimistic case. (c) Overhead ratio, pessimistic case. (d) Overhead ratio, optimistic case. (e) Average latency, pessimistic case. (f) Average latency, optimistic case. (g) Average hop count, pessimistic case. (h) Average hop count, optimistic case.

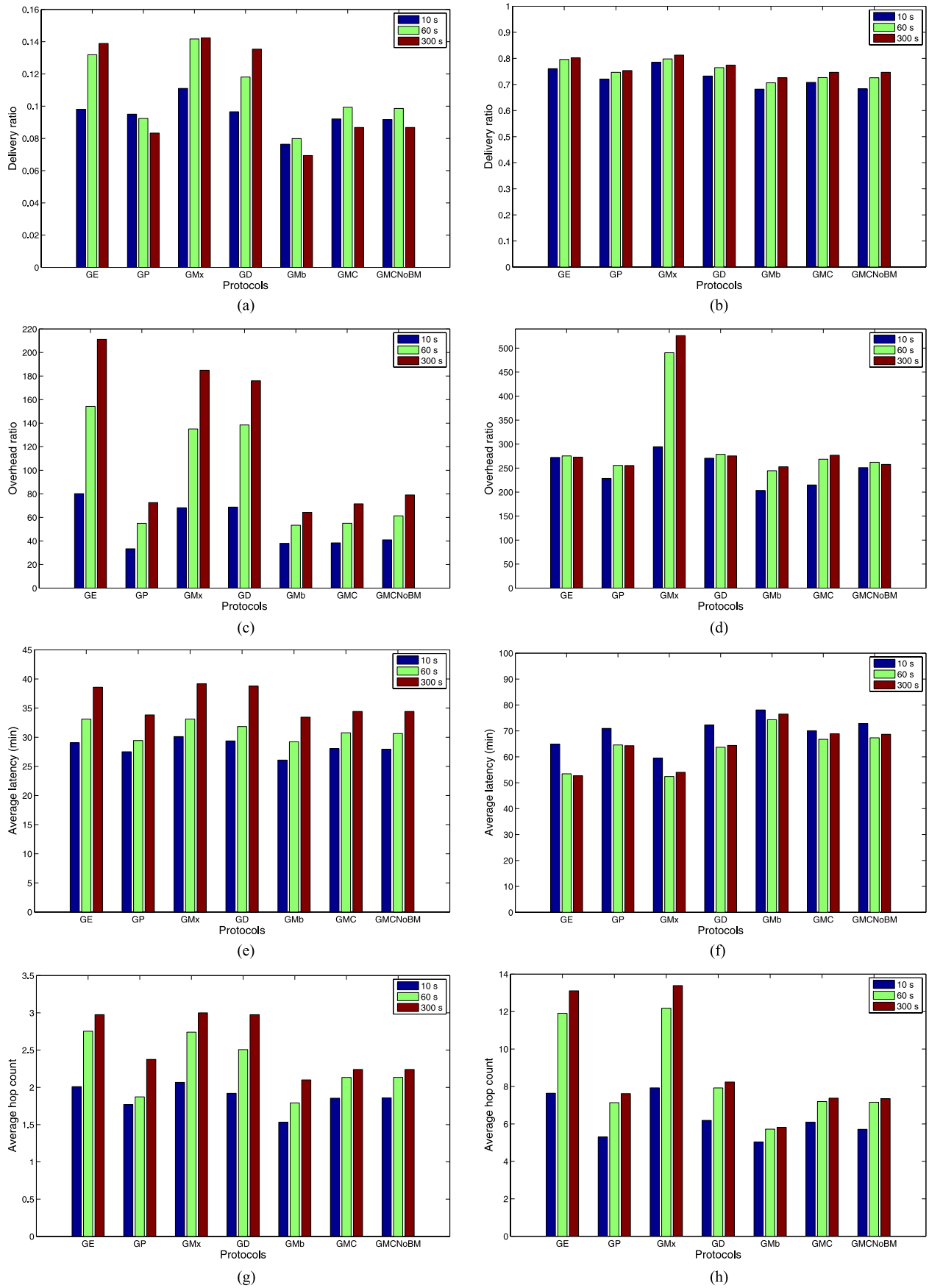


Fig. 7. Impact of the network traffic intensity. (a) Delivery ratio, pessimistic case. (b) Delivery ratio, optimistic case. (c) Overhead ratio, pessimistic case. (d) Overhead ratio, optimistic case. (e) Average latency, pessimistic case. (f) Average latency, optimistic case. (g) Average hop count, pessimistic case. (h) Average hop count, optimistic case.

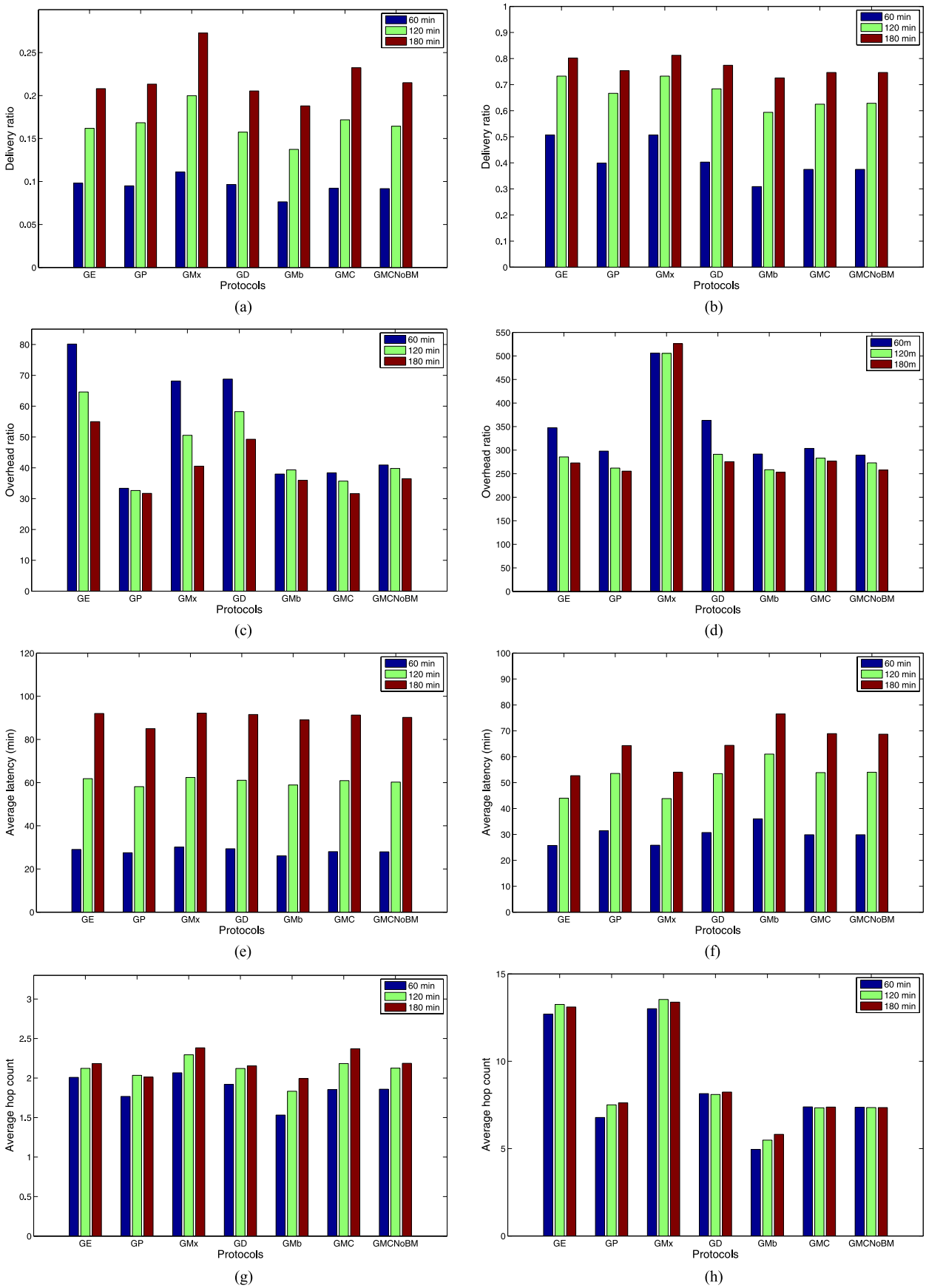


Fig. 8. Impact of the message TTL. (a) Delivery ratio, pessimistic case. (b) Delivery ratio, optimistic case. (c) Overhead ratio, pessimistic case. (d) Overhead ratio, optimistic case. (e) Average latency, pessimistic case. (f) Average latency, optimistic case. (g) Average hop count, pessimistic case. (h) Average hop count, optimistic case.

the redundant messages (i.e., the messages with a low delivery likelihood and the extra copies of the delivered messages) in the system. This advantage is more apparent for the optimistic cases: Without removing those redundant messages, the number of total message relays easily gets high, leading to a high overhead. By prioritizing the transmission queue, it also helps to utilize the transmission opportunities more effectively.

VI. Conclusion

In this paper, an advanced mobility-contact-aware geocast (GeoMobCon) scheme is proposed for large-scale urban VANETs. This new scheme has many new features compared with the existing geocast schemes: First, it is designed from a DTN point of view, enabling it to deal with the high mobility and transient connectivity in VANETs; second, different levels and aspects of vehicle mobility information and node-to-destination contact history information are employed, making GeoMobCon simple, scalable, and communication and computation effective; third, practical issues are well considered by introducing real-world trace analysis, trace-driven simulation, and efficient buffer management. Extensive performance comparisons with other protocols have shown the great advantages mentioned above.

Appendix

We can also plot the simulation results listed in Tables III–V. Figs. 6–8 show the impact of the communication transmission range, network traffic, and message TTL, respectively.

References

[1] Y. Ko and N. Vaidya, "GeoTORA: A protocol for geocasting in mobile ad hoc networks," in *Proc. IEEE ICNP*, 2000, pp. 240–250.

[2] W. Liao, Y. Tseng, K. Lo, and J. Sheu, "GeoGRID: A geocasting protocol for mobile ad hoc networks based on grid," *J. Internet Technol.*, vol. 1, no. 2, pp. 23–32, Dec. 2000.

[3] Y. Ko and N. Vaidya, "Flooding-based geocasting protocols for mobile ad hoc networks," *Mobile Netw. Appl.*, vol. 7, no. 6, pp. 471–480, Dec. 2002.

[4] T. Imieliński and J. Navas, "GPS-based geographic addressing, routing, and resource discovery," *Commun. ACM*, vol. 42, no. 4, pp. 86–92, Apr. 1999.

[5] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Efficient routing in intermittently connected mobile networks: The multiple-copy case," *IEEE/ACM Trans. Netw.*, vol. 16, no. 1, pp. 77–90, Feb. 2008.

[6] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic routing in intermittently connected networks," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 7, no. 3, pp. 19–20, Jul. 2003.

[7] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic routing in intermittently connected networks," in *Service Assurance with Partial and Intermittent Resources*, vol. 3126. Berlin, Germany: Springer-Verlag, 2004, pp. 239–254.

[8] L. Zhang, B. Yu, and J. Pan, "GeoMob: A mobility-aware geocast scheme in metropolitans via taxicabs and buses," in *Proc. IEEE INFOCOM*, 2014, pp. 1779–1787.

[9] L. Zhang, M. Ahmadi, J. Pan, and L. Chang, "Metropolitan-scale taxicab mobility modeling," in *Proc. IEEE GLOBECOM*, 2012, pp. 5404–5409.

[10] J. Burgess, B. Gallagher, D. Jensen, and B. Levine, "MaxProp: Routing for vehicle-based disruption-tolerant networks," in *Proc. IEEE INFOCOM*, 2006, pp. 1–11.

[11] G. Finn, "Routing and addressing problems in large metropolitan-scale internetworks," Adv. Res. Proj. Agency, Arlington, VA, USA, ISI/IRR-87180, DTIC Document, 1987.

[12] I. Stojmenovic, A. Ruhil, and D. Lobiyal, "Voronoi diagram and convex hull based geocasting and routing in wireless networks," *Wireless Commun. Mobile Comput.*, vol. 6, no. 2, pp. 247–258, Mar. 2006.

[13] I. Leontiadis and C. Mascolo, "GeOpps: Geographical opportunistic routing for vehicular networks," in *Proc. IEEE WoWMoM*, 2007, pp. 1–6.

[14] P. Cheng, K. Lee, M. Geria, and J. Härri, "GeoDTN+Nav: Geographic DTN routing with navigator prediction for urban vehicular environments," *Mobile Netw. Appl.*, vol. 15, no. 1, pp. 61–82, Feb. 2010.

[15] R. Raw and S. Das, "Performance analysis of P-GEDIR protocol for vehicular ad hoc network in urban traffic environments," *Wireless Pers. Commun.*, vol. 68, no. 1, pp. 65–78, Jan. 2013.

[16] S. Tsiachris, G. Koltsidas, and F. Pavlidou, "Junction-based geographic routing algorithm for vehicular ad hoc networks," *Wireless Pers. Commun.*, vol. 71, no. 2, pp. 955–973, Jul. 2013.

[17] C. Li, C. Zhao, L. Zhu, H. Lin, and J. Li, "Geographic routing protocol for vehicular ad hoc networks in city scenarios: A proposal and analysis," *Int. J. Commun. Syst.*, vol. 27, no. 12, pp. 4126–4143, Dec. 2014.

[18] Q. Yang, A. Lim, S. Li, J. Fang, and P. Agrawal, "ACAR: Adaptive connectivity aware routing for vehicular ad hoc networks in city scenarios," *Mobile Netw. Appl.*, vol. 15, no. 1, pp. 36–60, Feb. 2010.

[19] O. Kaiwartya, S. Kumar, and R. Kasana, "Traffic light based time stable geocast (T-TSG) routing for urban VANETs," in *Proc. IEEE IC3*, 2013, pp. 113–117.

[20] O. Kaiwartya, S. Kumar, D. K. Lobiyal, A. H. Abdullah, and A. N. Hassan, "Performance improvement in geographic routing for vehicular ad hoc networks," *Sensors*, vol. 14, no. 12, pp. 22 342–22371, Nov. 2014.

[21] C. Tripp-Barba *et al.*, "A multimetric, map-aware routing protocol for VANETs in urban areas," *Sensors*, vol. 14, no. 2, pp. 2199–2224, Jan. 2014.

[22] S. Cha, K. Lee, and H. Cho, "Grid-based predictive geographical routing for inter-vehicle communication in urban areas," *Int. J. Distrib. Sens. Netw.*, vol. 2012, 2012, Art. ID. 819 497.

[23] M. Di Felice, L. Bedogni, and L. Bononi, "Group communication on highways: An evaluation study of geocast protocols and applications," *Ad Hoc Netw.*, vol. 11, no. 3, pp. 818–832, May 2013.

[24] Y. Chen and Y. Lin, "A mobicast routing protocol with carry-and-forward in vehicular ad hoc networks," *Int. J. Commun. Syst.*, vol. 27, no. 10, pp. 1416–1440, Oct. 2012.

[25] S. Shivshankar and A. Jamalipour, "Spatio-temporal multicast grouping for content-based routing in vehicular networks: A distributed approach," *J. Netw. Comput. Appl.*, vol. 29, pp. 93–103, Mar. 2014.

[26] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Duke Univ., Durham, NC, USA, Tech. Rep. CS-200006, 2000.

[27] H. Zhu, S. Chang, M. Li, K. Naik, and S. Shen, "Exploiting temporal dependency for opportunistic forwarding in urban vehicular networks," in *Proc. IEEE INFOCOM*, 2011, pp. 2192–2200.

[28] H. Zhu *et al.*, "ZOOM: Scaling the mobility for fast opportunistic forwarding in vehicular networks," in *Proc. IEEE INFOCOM*, 2013, pp. 2832–2840.

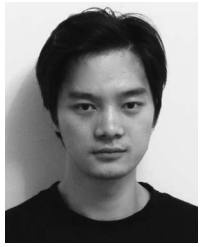
[29] R. Hall, "An improved geocast for mobile ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 2, pp. 254–266, Feb. 2011.

[30] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE simulator for DTN protocol evaluation," in *Proc. ICST SIMUtools*, 2009, p. 55.

[31] M. Amadeo, C. Campolo, and A. Molinaro, "Enhancing IEEE 802.11 p/WAVE to provide infotainment applications in VANETs," *Ad Hoc Netw.*, vol. 10, no. 2, pp. 253–269, Mar. 2012.

[32] H. Hartenstein and K. Laberteaux, "A tutorial survey on vehicular ad hoc networks," *IEEE Commun. Mag.*, vol. 46, no. 6, pp. 164–171, Jun. 2008.

[33] C. Lochert, B. Scheuermann, and M. Mauve, "Probabilistic aggregation for data dissemination in VANETs," in *Proc. ACM VANET*, 2007, pp. 1–8.

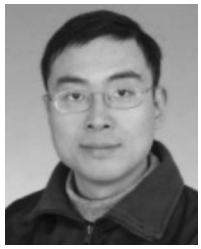


Lei Zhang (M'14) received the B.E. degree in information security from China University of Geosciences, Wuhan, China, in 2010 and the Ph.D. degree from the University of Victoria, Victoria, BC, Canada, in 2015.

He is currently with the Department of Computer Science, the University of Victoria. His research interests mainly reside in advanced wireless networks, including user mobility modeling, social characteristic, and security and privacy concerns.

Boyang Yu (S'13) received the B.S. and M.S. degrees in computer science from Nankai University, Tianjin, China, in 2006 and 2009, respectively. He is currently working toward the Ph.D. degree with the Department of Computer Science, University of Victoria, Victoria, BC, Canada.

His research interests include cloud computing, distributed systems, data-intensive services, and delay-tolerant networks.



Jianping Pan (SM'08) received the B.S. and Ph.D. degrees in computer science from Southeast University, Nanjing, China.

He did his postdoctoral research with the University of Waterloo, Waterloo, ON, Canada. He was also with Fujitsu Laboratories and NTT Laboratories. He is currently a Professor of computer science with the University of Victoria, Victoria, BC, Canada. His area of specialization is computer networks and distributed systems, and his current research interests include protocols for advanced networking, performance analysis of networked systems, and applied network security.

Prof. Pan has been serving on the Technical Program Committees of major computer communications and networking conferences, including the IEEE Conference on Computer Communications, the IEEE International Conference on Communications, the IEEE Global Communications Conference (GLOBECOM), the IEEE Wireless Communications and Networking Conference, and the IEEE Consumer Communications and Networking Conference. He is the Ad Hoc and Sensor Networking Symposium Co-Chair for IEEE GLOBECOM 2012 and an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.