CCR: Capacity-Constrained Replication for Data Delivery in Vehicular Networks

Yuchen Wu[1], Yanmin Zhu[1,2,*], Hongzi Zhu[1], Bo Li[1,3]
[1]Department of Computer Science and Engineering, Shanghai Jiao Tong University
[2]Shanghai Key Lab of Scalable Computing and Systems
[3]Department of Computer Science and Engineering, Hong Kong University of Science and Technology
* Corresponding Author
[1]{eaufavor, yzhu, hongzi}@sjtu.edu.cn; [2]bli@cse.ust.hk

***Abstract*—Given the unique characteristics of vehicular networks, specifically, frequent communication unavailability and short encounter time, packet replication has been commonly used to facilitate data delivery. Replication enables multiple copies of the same packet to be forwarded towards the destination, which increases the chance of delivery to a target destination. However, this is achieved at the expense of consuming extra already scarce bandwidth resource in vehicular networks. Therefore, it is crucial to investigate the fundamental problem of exploiting constrained network capacity with packet replication. We make the first attempt in this work to address this challenging problem. We first conduct extensive empirical analysis using three large datasets of real vehicle GPS traces. We show that a replication scheme that either underestimates or overestimates the network capacity results in poor delivery performance. Based on the observation, we propose a Capacity-Constrained Replication scheme or CCR for data delivery in vehicular networks. The key idea is to explore the residual capacity for packet replication. We introduce an analytical model for characterizing the relationship among the number of replicated copies of a packet, replication limit and queue length. Based on this insight, we derive the rule for adaptive adjustment towards the optimal replication strategy. We then design a distributed algorithm to dictate how each vehicle can adaptively determine its replication strategy subject to the current network capacity. Extensive simulations based on real vehicle GPS traces show that our proposed CCR can significantly improve delivery ratio comparing with the state-of-the-art algorithms.**

***Keywords*—vehicular networks, network capacity, packet replication, data delivery, analytical model, trace-driven simulations.**

I. Introduction

Vehicular networks have recently received significant attention in the community, which has fostered a wide range of emerging applications, such as driving safety [1], intelligent transport services [2], and infrastructure monitoring [3].

Efficient data delivery is essential for such applications, in which it is highly desired that data can be delivered with high success rate and low delay. However, unique characteristics of vehicular networks present several great challenges for data delivery, specifically, high vehicle mobility, frequent network disconnection [4] and short encounter time [5].

To achieve efficient data delivery, *packet replication* has been recognized as an effective approach for data delivery in vehicular networks. As illustrated in Fig. 1, replication enables multiple copies of the same data carried by different vehicles to be forwarded towards a target destination simultaneously, which increases the performance of delivery in terms of both higher delivery ratio and shorter delivery delay.

There has been a variety of algorithms proposed based on replication, which can be roughly divided into two categories: (1) the first category of algorithms uses a fixed replication number for each packet, i.e., sets a limit on the number of copies for each packet. Spray-and-wait [6] is such an example; (2) the second category of algorithms replicates packets based on a computed metric. When certain conditions on the metric are satisfied, a packet is replicated and the resulting copies are independently forwarded in the network. RAPID [7], trajectory-based [8], inter-contact time [4] and path likelihood [9] are such examples. An extreme case in this category is the epidemic algorithm [10], which replicates a packet each time when a vehicle encounters another vehicle.

However, *data replication consumes extra already scarce bandwidth resource in vehicular networks*. On one hand, in-

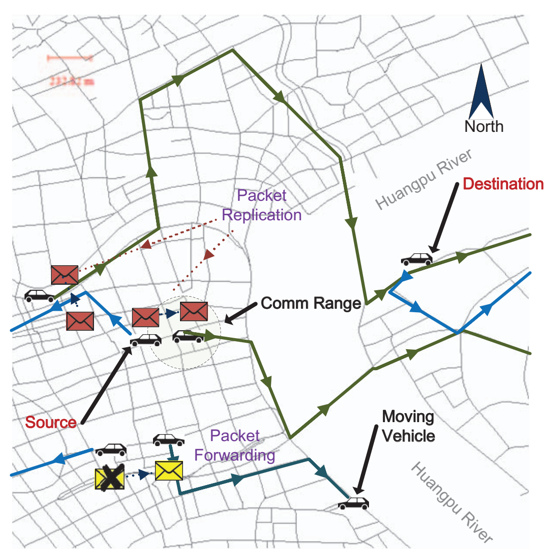

Figure 1. Illustration of packet forwarding and replication in vehicular networks. When a vehicle meets another, it can decide whether to replicate a packet or simply forward the packet to the relay node. Replication results in two copies being delivered towards to the destination. And with forwarding, the vehicle deletes the packet from its queue.

adequate replication might not obtain the desired performance gain; on the other hand, excessive replication can lead to poor utilization of the bandwidth, which inversely affects the performance. Therefore, the fundamental question for data replication in vehicle networks is to properly explore the network capacity without overloading. From algorithmic point of view, the problem becomes how to perceive the residual network capacity when a replication decision has to be made given the current network status. Our work in this paper attempts to address this problem.

In order to capture the network capacity, we first need to understand the effect of packet replication for data delivery in real-world vehicular networks. We conduct extensive empirical study with three large datasets of real vehicle GPS traces, i.e., 2,400 taxis in Shanghai, China, 1,600 buses in Shanghai, China, and 12,000 taxis in Shenzhen, China. There are two key observations from this study. On the one hand, it is demonstrated that the network capacity can be exploited for improving data delivery in vehicular networks, in which the delivery ratio can be significantly improved by 267% to 667%. On the other hand, a replication scheme that either underestimates or overestimates the network capacity results in suboptimal performance of data delivery.

In this paper, based on the above observations, we propose a Capacity-Constrained Replication scheme or CCR for data delivery in vehicular networks. *The key idea is to explore the residual network capacity for data replication*. There are two major challenges. First, it is widely known that it is difficult to obtain the accurate estimation of network capacity in a vehicular network with a number of uncontrollable factors involved such as the number of vehicles, vehicle mobility pattern, road topology, and traffic conditions. Second, the network traffic demand for data delivery can be changing over time, which is usually impossible to know in advance.

To tackle these challenges, we first introduce an analytical model for characterizing the relationship among replicated copies, replication limit and packet queue length. We use the *replication limit* to specify the maximum number of copies that a packet can replicate in the network. Based on this insight, we derive the rule for adaptive adjustment towards the optimal replication strategy. We then design a distributed algorithm to dictate how each vehicle can adaptively determine its replication strategy subject to the current network capacity. The salient feature in the proposed algorithm is that it only relies on the queue length observed, which can be easily measured and obtained. This makes it highly applicable in practice. Extensive simulations based on real vehicle GPS traces show that our proposed CCR can significantly improve delivery ratio comparing with the state-of-the-art algorithms.

We have made the main technical contributions in this paper as follows.

- We have conducted extensive empirical study with three large datasets of real vehicular GPS traces and reveal the potential and the problem of using packet predication in vehicular networks.
- Based on an analytical framework, we obtain the insight into the fundamental relationship among number of replicated copies, replication limit and packet queue length.
- We propose a fully distributed algorithm for adaptively determine its replication strategy subject to the current network capacity. *The main strength is that it removes the necessity of deriving the network capacity and only looks at queue lengths, which can be easily measured by individual vehicles.*
- We have performed extensive experiments based on large datasets of real vehicle GPS traces. Comparative study shows that CCR can improve data delivery ratio by 15% to 40% for static traffic demand and by 44% to 80% for dynamic traffic demand.

TABLE I
SUMMARY OF THREE VEHICLE TRACES

Traces	Shanghai Taxi	Shanghai Bus	Shenzhen Taxi
# of vehicles	2,400	1,600	12,000
Range (km $\times$ km)	133 $\times$ 69	133 $\times$ 69	27 $\times$ 97
Duration (hour)	17,280	17,280	720
Granularity (second)	60	60	60
Encounters (per hour)	352	891	425

The rest of the paper is organized as follows. The following section tries to understand the effect of packet replication with empirical study. In Section III, we propose an analytical framework for characterizing the relationship among replicated copies, replication limit and packet queue length. Section IV describes the design and implementation of CCR. Trace-driven simulations are presented in Section V. Section VI reviews related work. The paper concludes in Section VII.

II. EMPIRICAL STUDY

In this section, we aim at revealing the effect of packet replication on data delivery in vehicular networks with trace-driven empirical study.

A. Methodology of Trace-driven Empirical Study

For empirical study, we have conducted extensive simulations based on three large datasets of real vehicle GPS traces, as summarized in Table I.

We adopt the spray-and-wait algorithm [6] for packet replication and routing in vehicular networks, which uses a fixed replication limit. Suppose the replication limit is L. Then, an original packet is allowed to be replicated L times. As the source node encounters a vehicle, the packet is replicated on the vehicle. As a result, there are two copies in the network for the packet, each allowing to be replicated by $L/2$ times. The replication process proceeds until every copy of the packet allows no more replication.

The capacity of a vehicular network is dependent on a number of factors, such as number of vehicles, vehicle mobility, vehicle distribution, road topology and road traffic conditions. By using the traces, many of the factors have been reflected. Thus, these factors are fixed within the traces. To deliberately change the network capacity, we vary the number of vehicles in simulation. The traffic demand is another important system

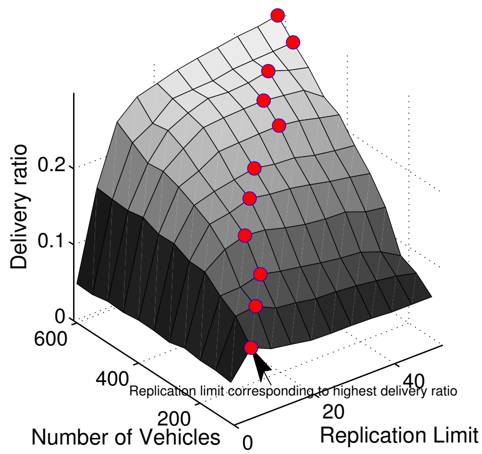

Figure 2. Delivery ratio vs. replication limit and number of vehicles (Shanghai Taxi).

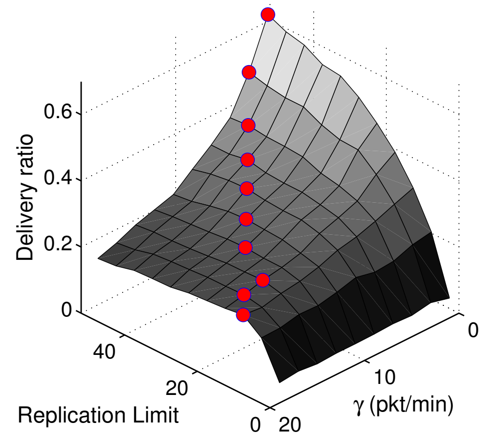

Figure 3. Delivery ratio vs. replication limit and packet generation rate γ (Shanghai Taxi).

parameter. To study the impact of traffic demand, we introduce a parameter called *packet generation rate (γ)* to control the traffic demand of a vehicle.

For each simulation, we randomly select 400 vehicles by default. The link bandwidth available for packet transmission each time two vehicles encounter is fixed. Thus, the number of packets that can be exchanged during one encounter between two vehicles is determined by the encounter duration.

We introduce a control parameter called *replication limit* for replication control, as defined as follows.

Definition 1 (Replication Limit)**.** *The replication limit for a packet is defined as the maximum number of copies that is allowed to be generated in the whole network.*

B. Performance Gain of Packet Replication

We first study the effect of packet replication on data delivery in vehicular networks. The delivery ratio as number of vehicles and replication limit are varied for Shanghai Taxi is shown in Fig. 2. Each red dot in the figure represents the highest delivery ratio among the simulations with different replication numbers and the same number of vehicles. When there are 600 vehicles, the delivery ratio is as low as 5% if no replication is used, i.e., replication limit is one. When a replication limit of 50 is used, the delivery ratio climbs to 30%. Thus, there is a 6x performance gain on delivery ratio when packet replication is used.

The delivery ratio as replication limit and packet generation rate are varied for Shanghai Taxi is shown in Fig. 3. We can find that when the packet generation rate is as low as one packet per minute, the performance gain of using packet replication is as high as 667%. In comparison, when the packet generation rate increases to 18 packets per minute, the performance gain decreases to 267%.

Then, we can make the first key observation as follows.

Key observation 1: Significant performance gain can be obtained by using packet replication to exploit the network capacity of a vehicular network. The performance gain of using packet replication is dependent on the capacity and the traffic demand of the whole network.

C. Impact of Replication Strategy

We next investigate the impact of replication strategy on data delivery. The replication strategy refers to the aggressiveness of generating replicated copies, which can be reflected by the parameter of replication limit. As shown in Fig. 2, when there are 200 vehicles in the network, the delivery ratio first increases as the replication limit is varied from one to 10 and then decreases as a large replication limit is used. This is because the network capacity of the vehicular network is always limited. As more replicated copies are generated, some of the limited communication opportunities are wasted by unnecessary transmissions of replicated copies. As a result, the delivery ratio decreases, other than increases, as more replicated copies are injected in the network. We have similar findings when there are different numbers of vehicles in the system. The difference is that the maximum delivery ratio becomes higher as there are more vehicles and the corresponding replication limit is larger. This is because the network capacity becomes larger when there are more vehicles and thus allows more replicated copies.

As shown in Fig. 3, as a higher packet generation rate is applied, the maximum delivery ratio decreases and the corresponding replication limit becomes smaller. This is because the total traffic demand becomes larger and the residual network capacity is smaller, hence allowing fewer replicated copies.

Thus, we can make the second key observation as follows.

Key observation 2: A replication scheme that either underestimates or overestimates the network capacity leads to poor delivery performance. In addition, for a vehicular network with given traffic demand, there exists an optimal replication limit that produces the maximum delivery ratio.

III. Characterizing Effect of Replication

Key observation 2 from the empirical study suggests that it is crucial to make appropriate control on packet replication

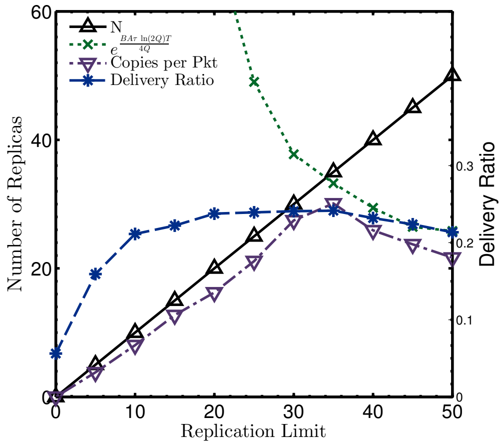

Figure 4. Relationship among number of replicas, replication limit, theoretical free growth (computed with measured queue length), along with measured number of replicas and delivery ratio.

for higher delivery ratio in vehicular networks. Then, we have to answer the important question: *what is the optimal replication strategy for different network configurations?* This section tries to answer this question. First, we give a simplified analytical model to facilitate analysis, then drive the dynamical formulation of number of replicas of a packet, and finally obtain the rule for determining the optimal replication limit.

A. Analytical Model

We assume all vehicles have the same communication range. Every two vehicles can transmit data when they are within the communication range. For analysis simplification, we assume the bandwidth of the link between two vehicles is B packets per unit time. Each of the two vehicles shares the link bandwidth equally. The encounter duration of the two vehicles is d. Then, the total amount of data can be exchanged for a vehicle during one encounter is $Bd/2$.

Every vehicle may generate packets. Data packets are of the same size and priority. Each packet has one single destination vehicle. Packets have the same TTL (Time-To-Live).

We assume the buffer of a vehicle is sufficiently large, so a vehicle can receive and save packets without deleting or replacing other packets. To simplify the analysis, we model the buffer as a cyclic first-in-first-out (FIFO) queue. The length of the queue on vehicle v is Q_v, which equals to the number of packets it currently holds. However, as mentioned before, the bandwidth is fixed. As a result, only a few packets in the queue can be sent within one encounter. A sent packet is removed from the head of queue and inserted to the rear.

A packet gets a new replica every time it is replicated at one encounter. The number of replicas (including the original) of packet p is denoted by $n_p(t)$ as it varies over time. Let N denote the replication limit. Then, the number of replicas for any packet is no more than N, i.e., $n_p(t) \leq N$.

B. Characterizing Dynamical Number of Replicas

It is highly desirable to derive how the delivery performance changes as the replication limit is varied. However, it is extremely difficult as the delivery performance is dependent on many factors and most of them may not be available for analysis, such as node mobility and road topology. To walk around this problem, we instead look at the number of replicas of a packet. It is intuitive that for a given packet, a larger number of replicas indicates a higher delivery performance.

The number of replicas of a packet is dynamic and changes over time, which is dependent on network capacity, traffic demand and replication limit. Unfortunately, in practice we do not know the network capacity and traffic demand. We find that the queue length can be an important factor. Thus, we derive how the number of replicas of a packet grows over time, in relation to packet queue length.

Our following analysis assumes a basic replication scheme: a packet is always replicated if it is sent from a vehicle to the encountered vehicle until its number of replicas reaches the replication limit, N. This gives us the first relationship between number of replica and replication limit:

$$n_p(t) \leq N, \forall p, t. \tag{1}$$

Then, we discuss how $n_p(t)$ grows over time. It should be emphasized that this analysis applies to a stable network, where queue length is not varying dramatically.

Let the total encounter duration of vehicle v during a short time period Δt be denoted by d_v. Then, the total number of packets sent in this duration is $Bd_v/2$. Let the average encounter duration be denoted by τ and the encounter times in Δt by a_v. Then, d_v equals $a_v \times \tau$.

Let $\mathcal{K}$ denote the set of vehicles carrying p at time t. Because the relative position of p in the queue of a vehicle $v \in \mathcal{K}$ is random, the maximum expected number of times, Δm_v, of p that the packet can be replicated in Δt is,

$$\Delta m_v = \frac{B a_v \tau}{2 Q_v}. \tag{2}$$

Thus, the total number of newly generated replicas of p, Δn_p, during Δt is,

$$\Delta n_p = \sum_{v \in \mathcal{K}} \Delta m_v = \sum_{v \in \mathcal{K}} \frac{B a_v \tau}{2 Q_v}. \tag{3}$$

The queue length Q_v of v is dependent on the previous encounters which is independent of a_v. To ease the analysis, we assume a uniform distribution for Q_v. However, we should note that this analysis can also be easily extended to other distributions for Q_v. For an arbitrary p, the expected increase rate of the number of replicas of p is

$$E \left[\frac{\Delta n}{\Delta t}\right] = E\left[\sum_{v \in \mathcal{K}} \frac{B a_v \tau}{2 Q_v \Delta t}\right] \tag{4}$$

$$= \frac{|\mathcal{K}|}{2} E\left[\frac{B a_v \tau}{\Delta t}\right] E\left[\frac{1}{Q_v}\right] = \frac{n_p(t)}{2} B A \tau E\left[\frac{1}{Q_v}\right] \tag{5}$$

$$\approx \frac{n_p(t)}{2} B A \tau \frac{\ln(2Q)}{2Q} = \frac{B A \tau \ln(2Q)}{4Q} n_p(t), \tag{6}$$

where A is the expected encounter frequency and Q is the expected queue length of the network.

With (6), we have the following differential equation for $n_p(t)$:

$$\begin{cases} \dfrac{dn_p(t)}{dt} = \dfrac{BA\tau\ln(2Q)}{4Q} n_p(t) \\ n_p(0) = 1 \end{cases}. \quad (7)$$

By solving (7), we have

$$n_p(t) = e^{\frac{BA\tau\ln(2Q)}{4Q}t}. \quad (8)$$

This gives us the free growth of packet replicas. From (8), we know *the number of replica grows faster when vehicles encounter each other more often and it slows down if packet queues are very long*.

The number of replicas is actually constrained by both the two factors: *free growth* as Eq. 8 and *replication limit* as Eq. 1.

$$n_p(t) = \min(N, e^{\frac{BA\tau\ln(2Q)}{4Q}t}). \quad (9)$$

C. Determining Optimal Replication Limit

Since for a given packet, its life cycle is constrained by its TTL. We thus focus on the question: *when can n_p reach its maximum n_p^* before the TTL expires and what is the corresponding optimum replication limit N^*?*

Importantly, these two factors of free growth and replication limit for $n_p(t)$ are not independent from each other. A larger limit gives replicas more space to grow. As a consequence, however, the queue length becomes longer, which slows down the growth speed. A shorter queue length allows replicas to grow faster but it requires a smaller replication limit. This observation indicates that an optimal replication number must consider both growth speed and replication limit.

Theorem 1. *For a given duration T, the maximum of $n_p(T)$ is reached when the following equation satisfied.*

$$N = e^{\frac{BA\tau\ln(2Q)}{4Q}T}. \quad (10)$$

Proof: For a given duration T, when we increase N, Q increases as more copies of every packet are allowed to be generated. So, the free growth in this duration, $e^{\frac{BA\tau\ln(2Q)}{4Q}T}$, is a monotonically decreasing function of N. Considering Eq. 9, as N is monotonically increasing function of itself, the maximum value is reached when the replication limit and the free growth equals. This concludes the proof. ■

Unfortunately, Eq. 10 does not give us the analytical result of the optimum N^* as we do not have the analytical relationship between N and Q. In a practical vehicular network, the queue length depends on many factors including replication limit, encounters, packet generation and extinction. It is hard to formulate the relationship.

To illustrate how the number of replicas of packet p varies with the replication limit. We conduct simulation-based empirical study with the Shanghai taxi traces to derive the relationship between the number of replicas with the TTL and the replication limit. Fig. 4 shows this relationship, along with the delivery ratio as the replication limit varies. From the figure, we can easily find that the total number of generated replicas reaches the maximum as the growth equals N. This verifies our analytical result in Theorem 1. In addition, we find that the delivery ratio is maximized as the total number of replicas reaches the maximum. This confirms our intuition.

Therefore, although Theorem 1 does not lead to an analytical result for the optimal N^*,it does give us a rule for adjustment towards the optimum. We know the current N and measure the current queue length, Q. Then we can compute the free growth of packet replicas, $e^{\frac{BA\tau\ln(2Q)}{4Q}TTL}$. We can see from Fig. 4 that if the free growth is smaller than N, then N should be decreased; otherwise, it should be increased. This rule will guide our design of the distributed algorithm.

IV. Design and Implementation

In this section, we describe the design and implementation of our distributed algorithm for each vehicle to dynamically determine its replication limit and make packet replication decisions upon encounters.

A. Overview

The distributed algorithm runs by each individual vehicle, which consists of three main components: 1) *packet queue length estimation*, 2) *replication limit adjustment* and 3) *replication control*. The first component estimates the average queue length of vehicles by measuring the queue length of each vehicle that has been encountered. This serves as the basis for adjustment of replication limit. The second component adjusts the current replication limit according to the estimated average queue length. Note that each vehicle maintains a replication limit for each newly packet generated to follow. The third component makes the decision on which packet to replicate on an encountered vehicle.

Our algorithm needs some information about encountered vehicles, including packet queue length, encounter frequency, and packet IDs in queue. The algorithm requires each vehicle to exchange such information upon encountering with each other. When three or more vehicles are in the communication range of each other, we rely on the media contention mechanism of the link layer to address the link scheduling issue.

B. Estimating Queue Length

As the expected packet queue length of the whole network is important to the adjustment of replication limit, each node must estimate this value by measuring the queue lengths of all encountered vehicles.

The packet queue length of a vehicle can be very different from each other, dependent on its encounter behaviors. In addition, the packet queue length is also varying over time, reflecting the current network capacity, traffic demand and replication strategy.

To estimate the queue length, we adopt the moving average technique. Each node v maintains an estimated network queue length Q_v. Let q_v denote the length of the packet queue at v.

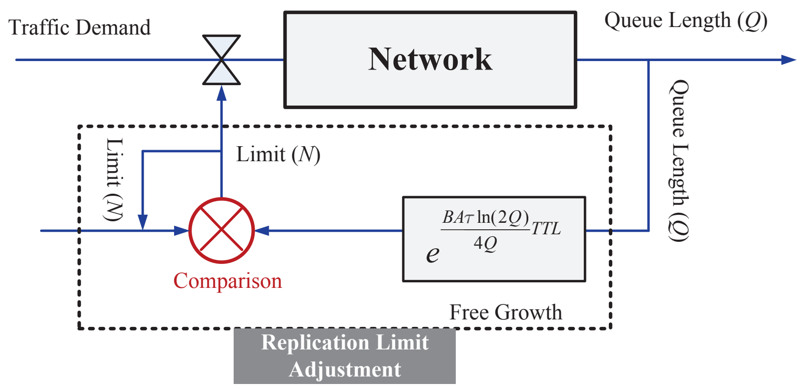

Figure 5. The architecture of the feedback-based adjustment of replication limit. Each vehicle keeps measuring queue lengths of encountered vehicles and estimating the queue length for the whole network. Then it compares the limit and the free growth. Based on this comparison result, the vehicle adjusts its replication limit.

Algorithm 1: *ReplicationControl*

Input: $\mathbb{Q}$: The packet queue, N

while in encounter duration and $\mathbb{Q}$ is not empty do
    $p \leftarrow \mathbb{Q}$.POP();
    $r_p \leftarrow \max(1, r_p \times \frac{N}{N_p})$;
    $N_p \leftarrow N$;
    if $r_p \geq 2$ then
        $r_p \leftarrow r_p/2$;
        $p' \leftarrow$ COPY(p);
        SEND(p');
        $\mathbb{Q}$.PUSH(p);
    else
        FORWARD(p);
    end
end

As vehicle v encounters u, it updates its estimate of the packet queue length as follows,

$$Q'_v = \alpha \times \frac{q_v + q_u}{2} + (1 - \alpha) \times \frac{Q_v + Q_u}{2}, \tag{11}$$

where α is a constant decay factor between 0 and 1.

The decay factor is an important design parameter and may impact the data delivery performance. A larger decay factor makes the estimate of the packet queue length more sensitive to the variation of queue length. A smaller factor makes the estimation more stable but cannot quickly adapt to the change of network. We will study the impact of the factor and discuss the selection of the factor in the next section.

C. Adjusting Replication Limit

With the estimated queue length of the network, we then develop an adjustment scheme for the replication limit maintained by each vehicle. The main rule for adjustment is based on Theorem 1.

The main rationale for this adjustment is feedback-based control, as illustrated in Fig. 5. The network can be considered as a system under control. The network takes traffic demand and the replication limit as inputs and generates the packet queue length as output. The replication limit adjustment component is the control unit based on the feedback of queue length. It compares the current replication limit N, and the theoretical free growth $e^{\frac{BA\tau \ln(2Q)}{4Q}TTL}$ with the estimated Q. If N is larger than the free growth, the replication limit is decreased; otherwise, it is increased.

The adjustment step is the key issue. It is desirable to approach N^* as soon as possible. We adopt the Halley's method for numerically calculating the Lambert W function which is defined by $z = W(z)e^{W(z)}$ [11]. It is similar to Newton's method, which determines the direction of search and the step size to approach the target value. Let e^{β} denote the free growth, the detailed adjustment is shown as follows:

$$\frac{1}{N'} = \frac{1}{N} - \frac{\frac{e^{\beta}}{N} - 1}{e^{\beta}(\frac{1}{N} + 1) - \frac{(\frac{1}{N}+2)(\frac{e^{\beta}}{N}-1)}{\frac{2}{N}+2}}. \tag{12}$$

D. Replication Control

We finally describe the replication control. The basic idea of the distributed replication control is to replicate packets as soon as possible from the source node to other vehicles. But, the number of copies is smaller than the replication limit assigned to the packet, N_p. After the total number of copies reaches N_p, each of the copies is forwarded towards the destination, where many of the routing algorithms for vehicular network can be employed. Thus, it should be stressed that our approach is complimentary to existing routing algorithms.

To better adapt to the changing network capacity, it is beneficial to allow a packet in flight to dynamically change its replication limit. The key issue is that when the packet was generated it had been assigned a replication limit. Then, derivative copies are allocated with different portions of the assigned replication limit. Thus, the change to the existing replication limit must be carried out in a distributed fashion. To facilitate this distributed adjustment on both the source node and the relaying nodes, each copy maintains two values:

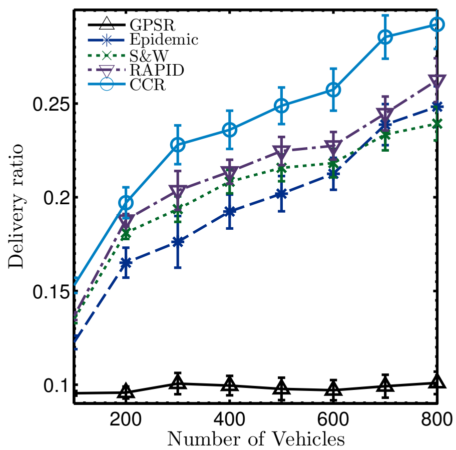

Figure 6. Delivery ratio vs. number of vehicles (Shanghai Taxi).

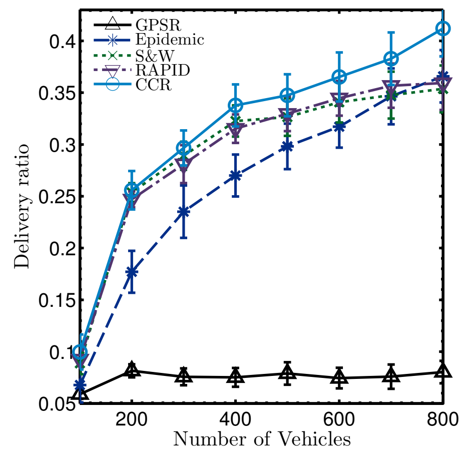

Figure 7. Delivery ratio vs. number of vehicles (Shenzhen Taxi).

its replication limit, N_p, and the remaining budget, r_p. If the relaying node u has a different replication limit N_u, it proportionally changes r_p to $r_p \frac{N_u}{N_p}$, and replaces N_p with N_u. The details are shown in Algorithm 1.

V. Performance Evaluation

In this section we first present the methodology and experimental results, then introduce the compared algorithms and finally present the evaluation results.

A. Methodology and Experimental Setup

To evaluate the performance of our approach CCR, we have conducted extensive simulations based on the three datasets of real vehicle GPS traces, as introduced in Section II.

In simulation, the following default settings are used. 400 vehicles are randomly selected from one of the trace datasets. The packet generation rate is 10 packets per minute. The bandwidth is set to allow the transmission of one packet per second. The communication range is 300 meters. For CCR, the average encounter duration, τ, is 20 seconds, TTL is one hour, and the decay factor α is 0.35.

We use one week of the Shanghai taxis from August 30 to September 5, 2007, one week of Shenzhen taxis from September 6 to September 12, 2009, and three days of Shanghai buses from October 15 to October 17, 2006.

B. Compared Algorithms

We compare CCR with the following routing algorithms representative for vehicular networks.

- **GPSR** [12]. It is a single copy routing algorithm, which forwards packets to relays closer to the destinations.
- **Epidemic** [10]. It replicates packets as many as possible to every other vehicle it encounters.
- **Spray-and-wait** [6]. It determines a fixed replication limit, which is 10%-15% of the number of nodes
- **RAPID** [7]. It decides whether to replicate a packet based on the probability of encountering the destination.

C. Impact of Number of Vehicles

We study the performance of the algorithms when the number of vehicles is varied. The comparisons using the three trace datasets are shown in Fig. 6, Fig. 7 and Fig. 8, respectively.

We can find that CCR substantially outperforms the four algorithms for all traces. Compared with epidemic, the improvement of delivery ratio of CCR is as large as 25% for shanghai taxis, 47% for Shenzhen taxis, 40% for Shanghai buses. Compared with Spray-and-wait and RAPID, the improvement is as large as 15%, 20% and 10% for the three datasets, respectively.

The delivery ratios become larger with the increasing number of vehicles except that of GPSR. This is because when the network has more vehicles, there are more possible paths. Replication-based algorithms take advantage of it and produce higher delivery ratios. However, GPSR, a single copy routing algorithm, fails to benefit from it.

D. Impact of Network Traffic

We first investigate the impact of constant network traffic demand, where the packet generation rate is fixed during the whole simulation. The delivery ratio as the packet generation rate is varied for all algorithms is plotted in Fig. 9. CCR has a delivery ratio almost as high as epidemic when the traffic demand is low, and outperforms all other algorithms when the traffic demand is high. When the demand is low, epidemic and CCR leverage the network capacity for data delivery. When the demand is high, the delivery ratio of the epidemic algorithm quickly degrades, which use blind replication strategy. In comparison, CCR can adapt to the residual network capacity and make appropriate replication control, and thus it is constantly better than all other algorithms. The delivery ratios of all replication-based algorithms decrease because when there are more packets in the network, the delivery opportunity of each packet is smaller since the network capacity is limited.

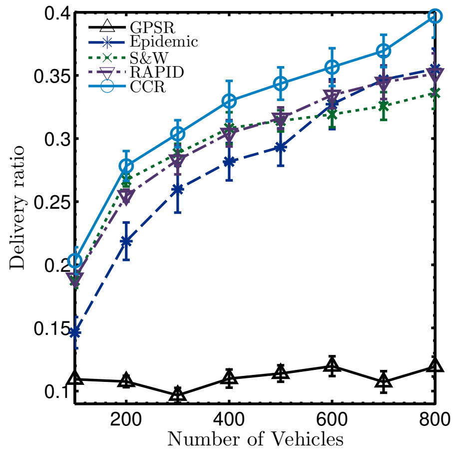

Figure 8. Delivery ratio vs. number of vehicles (Shanghai Bus).

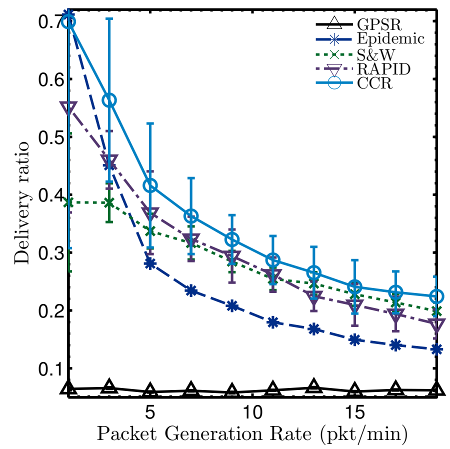

Figure 9. Delivery ratio vs. packet generation rate γ (Shanghai Taxi).

VI. Related Work

In this section we give an overview of related work. First, we review routing algorithms for vehicular networks, and then discuss existing studies on network capacity of wireless networks.

A. Routing Algorithms in Vehicular Networks

A number of routing algorithms have been proposed for vehicular networks, which can be divided into several categories.

One category of them maintains at most one copy for each packet. With the knowledge of node locations, GPSR [12], CAR [13] and GeOpps [14] forward packets to relays which are closer to their destinations. TBD [15] and TSF [16] forward packet along roads, which have the minimal delay according to traffic information. When all the contacts between vehicles are known, the routing problem can be formulated and solved by linear programming [17]. As the complete contacts cannot be foreknown in practice, estimation of Euclidean distance [18] or delay [19] can be used for path selection.

With the observation that forwarding multiple copies of a packet through different paths usually reduces its delay and increases the delivery probability, packet replication has been adopted in two other categories of routing algorithms for vehicular networks.

The second category uses a fixed number of replicas. Spray-and-wait [6] tries to answer the question how many copies are enough to achieve certain delay constrain. Following the basic mobility model of exponentially distributed inter-meeting time, the expected delay of each copy can be calculated. The delay bound of spraying a certain number of copies can be derived from these values. This answers the question above. Source-spray and binary-spray, two different approaches of spraying copies, are proposed in this paper as well. To further reduce the number of copies for cost efficiency, a multiple period spraying scheme [20] tries not to send all the copies once, but to add copies in each period according to the packet's urgency. R3 [21] studies unified metrics for path across networks with diverse connectivity such as meshes, MANETs and DTNs. It introduces replication gain for delay if multiple paths are selected. The routing algorithm of R3 selects the best two paths for each packet.

The third category decides whether to copy packets whenever two vehicles encounter each other. RAPID [7] transfers routing metric about path into utilities for each packet. Thus, a resource allocation problem is formulated in this work when packets have their own cost and benefits. By solving this problem, packets are copied to those relays which increase the benefits the most. RAPID also adopts the exponential distribution of encounter probability. Later works use more sophisticated tools such as Markov chain to model encounter between nodes. By predicting node movement [8] and meeting time [4], more accurate metrics are achieved which lead to better performance. Other issues such as energy and storage saving are discussed in [22].

Little existing work has considered the problem of using appropriate packet replication to exploit the residual network capacity. In this work, we have proposed an approach to adaptively determine the replication strategy subject to the current residual network capacity. In addition, our work is complementary to most of the existing routing algorithms for vehicular networks by adding replication control.

B. Network Capacity of Wireless Networks

The capacity of wireless networks has attracted significant research interests. In a static wireless network, the total capacity of n optimally located nodes is $O(\sqrt{n})$ and per-node throughput scales as $O(1/\sqrt{n})$, as shown in [23]. This result implies that static wireless networks are not scalable. However, when considering the mobility of nodes, the capacity is largely increased. The throughput of per source-destination pair remains constant [24] when total number of nodes increases. The question whether delay keeps low when mobility increases capacity answers in [25]. This work shows that mobile ad-hoc networks can provide guarantees on the delay if the patterns in the mobility of nodes are exploited. Further studies of delay and capacity establish the trade-off: $delay/rate \geq O(n)$ [26].

This result indicates that throughput is the cost to reduce delay. Several mobility models are studied in [27]. Using a unified framework based on those models, this work demonstrates that there is a critical delay and the delay capacity trade-off exists only when it is greater than the critical value. As mentioned in [27], these theoretical studies on capacity use idealized models, such as unit disk model, i.i.d. model, random way-point model, Brownian model, and random walk model. Vehicle nodes have very different mobility behaviors and thus existing results can hardly be applied to vehicular networks. A few studies [28] [29] attempt to characterize the capacity of a vehicular network.

In summary, most of these studies assume unrealistic road topology, node mobility, and vehicle distribution. These capacity results can hardly be used for replication control in a real vehicular network.

VII. Conclusion

Packet replication has been recognized as an effective approach for data delivery in vehicular networks and many replication-based algorithms have been proposed. However, the main problem of using packet replication is not to overwhelm the network and not to leave network capacity underutilized. In response to this crucial problem, we propose CCR, a capacity-constrained replication scheme for data delivery in vehicular networks. We theoretically derive the important relationship among replicated copies, replication limit and queue length, and obtain the rule for determining the appropriate replication limit. A distributed algorithm is then designed, which adaptively determine the replication strategy that is suitable for the currently residual network capacity. Extensive simulations based on the large datasets of real vehicular GPS traces show that our approach CCR significantly improves delivery ratio comparing with the state-of-the-art algorithms.

Acknowledgements

This research is supported in part by Shanghai Pu Jiang Talents Program (10PJ1405800), Shanghai Chen Guang Program (10CG11), NSFC (No. 61170238, 60903190, 61027009, 61202375, 61170237), MIIT of China (2009ZX03006-001-01), Doctoral Fund of Ministry of Education of China (20100073120021), National 863 Program (2009AA012201 and 2011AA010500), HP IRP (CW267311), SJTU SMC Project (201120), STCSM (08dz1501600, 12ZR1414900), Singapore NRF (CREATE E2S2), NSFC/RGC (N-HKUST610/11), HKUST (RPC11EG29, SRFI11EG17-C, SBI09/10.EG01-C), Guangdong Bureau of Science and Technology(GDST11EG06), China Cache Corp. (CCNT12EG01), and Program for Changjiang Scholars and Innovative Research Team in Universities of China (IRT1158, PCSIRT).

References

[1] F. Farnoud and S. Valaee, "Reliable broadcast of safety messages in vehicular ad hoc networks," in *Proc. IEEE INFOCOM*, 2009.

[2] H. Hartenstein, B. Bochow, A. Ebner, M. Lott, M. Radimirsch, and D. Vollmer, "Position-aware ad hoc wireless networks for inter-vehicle communications: the fleetnet project," in *Proc. ACM MOBIHOC*, 2001.

[3] P. B. Gibbons, B. Karp, Y. Ke, S. Nath, and S. Seshan, "Irisnet: An architecture for a worldwide sensor web," *IEEE Pervasive Computing*, vol. 2, no. 4, pp. 22–33, 2003.

[4] H. Zhu, S. Chang, M. Li, S. Naik, and S. Shen, "Exploiting temporal dependency for opportunistic forwarding in urban vehicular network," in *Proc. IEEE INFOCOM*, 2011.

[5] V. Bychkovsky, B. Hull, A. K. Miu, H. Balakrishnan, and S. Madden, "A measurement study of vehicular internet access using in situ Wi-Fi networks," in *Proc. ACM MOBICOM*, 2006.

[6] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Efficient routing in intermittently connected mobile networks: The multiple-copy case," *IEEE/ACM Transactions on Networking*, vol. 16, pp. 77 – 90, 2008.

[7] A. Balasubramanian, B. N. Levine, and A. Venkataramani, "Replication routing in DTNs: A resource allocation approach," *IEEE/ACM Transactions on Networking*, vol. 18, pp. 596 – 609, 2010.

[8] Y. Wu, Y. Zhu, and B. Li, "Trajectory improves data delivery in vehicular networks," in *Proc. IEEE INFOCOM*, 2011.

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

[10] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Technical Report CS-2000-06, Department of Computer Science, Duke University, Tech. Rep., 2000.

[11] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the lambert w function," *Advances in Computational Mathematics*, vol. 5, pp. 329–359, 1996.

[12] B. Karp and H. T. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," in *Proc. ACM MOBICOM*, 2000.

[13] V. Naumov and T. R. Gross, "Connectivity-aware routing (car) in vehicular ad-hoc networks," in *Proc. IEEE INFOCOM*, 2007.

[14] I. Leontiadis and C. Mascolo, "Geopps: Opportunistic geographical routing for vehicular networks," in *Proc. IEEE WOWMOM workshop on Autonomic and Opportunistic Communications*, 2007.

[15] J. Jeong, S. Guo, Y. Gu, T. He, and D. Du, "TBD: Trajectory-based data forwarding for light-traffic vehicular networks," in *Proc. IEEE ICDCS*, 2009.

[16] ——, "TSF: Trajectory-based statistical forwarding for infrastructure-to-vehicle data delivery in vehicular networks," in *Proc. IEEE ICDCS*, 2010.

[17] S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," in *Proc. ACM SIGCOMM*, 2004.

[18] J. Leguay, T. Friedman, and V. Conan, "Evaluating mobility pattern space routing for DTNs," in *Proc. IEEE INFOCOM*, 2006.

[19] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Efficient routing in intermittently connected mobile networks: The single-copy case," *IEEE/ACM Transactions on Networking*, vol. 16, pp. 63–76, 2008.

[20] E. Bulut, Z. Wang, and B. K. Szymanski, "Cost-effective multi-period spraying for routing in delay-tolerant networks," *IEEE/ACM Transactions on Networking*, vol. 18, pp. 1530 – 1543, 2010.

[21] X. Tie, A. Venkataramani, and A. Balasubramanian, "R3: Robust replication routing in wireless networks with diverse connectivity characteristics," in *Proc. ACM MOBICOM*, 2011.

[22] T. Small and Z. J. Haas, "Resource and performance tradeoffs in delay-tolerant wireless networks," in *ACM SIGCOMM workshop on Delay-tolerant networking*, 2005.

[23] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, pp. 388 – 404, 2000.

[24] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, pp. 477 – 486, 2002.

[25] N. Bansal and Z. Liu, "Capacity, delay and mobility in wireless ad-hoc networks," in *Proc. IEEE INFOCOM*, 2003.

[26] M. J. Neely and E. Modiano, "Capacity and delay tradeoffs for ad hoc mobile networks," *IEEE Transactions on Information Theory*, vol. 51, pp. 1917–1937, 2005.

[27] G. Sharma, R. Mazumdar, and N. B. Shroff, "Delay and capacity trade-offs in mobile ad hoc networks: a global perspective," *IEEE/ACM Transactions on Networking*, vol. 15, pp. 981 – 992, 2007.

[28] F. E. Ali, B. Ducourthial, and S.-M. Senouci, "On the capacity of a linear vehicular network," in *Proc. IEEE VTC*, 2011.

[29] M. Nekoui and M. Nekoui, "Vanet capacity scaling under the physical model," in *Proc. Conference on Information Sciences and Systems*, 2008.