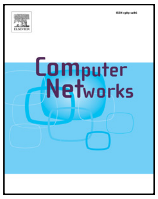
Computer Networks 101 (2016) 90–103

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet



ECDS: Efficient collaborative downloading scheme for popular content distribution in urban vehicular networks

Wei Huang[a,b], Liangmin Wang[a,c,*]

[a] School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China
[b] Portal Search Research, Baidu (China) Co., Ltd, Beijing 100085, China
[c] Co-Innovation Center for Information Supply & Assurance Technology, Anhui University, Hefei 230601, China

ARTICLE INFO

Article history:
Received 30 July 2015
Revised 17 January 2016
Accepted 2 February 2016
Available online 10 February 2016

Keywords:
Popular content distribution
Collaborative downloading
Relay selection
Generation selection
Scale-free property

ABSTRACT

The recent development of the Vehicular Ad-hoc Networks (VANETs) has motivated an increasing interest in in-vehicle consumption, and hence, the Popular Content Distribution (PCD) has become a heated issue. Compared with PCD solutions based on the widely-used cellular networks and Dedicated Short Range Communications (DSRC), solutions based on Collaborative Downloading (CD) are more economical and efficient. Due to the limited bandwidth, the On-Board Units (OBUs) passing through a Road Side Unit (RSU) can only download a portion of the popular content. To get over that drawback and to effect a collaborative downloading, a P2P network should be constructed among the OBUs which fall out of the RSUs coverage. In this paper, we address the efficient collaborative downloading scheme (ECDS) for PCD in urban traffic scenarios. To adapt to the rapid-changing characteristics of the VANET topology, a new cell-based clustering scheme is proposed, which greatly simplifies the modeling. Besides a strategy of inter-cluster Relay Selection is proposed to construct a pear-to-pear (P2P) network of scale-free property, which will help enhancing the information spread. Furthermore, another inter-cluster strategy of generation selection is to be collaborated to accelerate the dissemination process in the P2P network. The comparison experiments to two up-to-date collaborative PCD protocols demonstrate the high performance of the proposed scheme, i.e. ECDS.

1. Introduction

Vehicular Adhoc NETworks (VANETs), which provide high-rate data communication between the *On-Board Units* (OBUs) and *Road Side Units* (RSUs), are among the most successful industrial applications of the *Internet of Things* (IoT) [1]. RSUs are deployed along the roads so that OBUs may request any internet-based service (e.g. multimedia, TV shows, games, real-time traffic concerns, etc.). Recent development and standardization in VANETs have aroused commercial interest in in-vehicle consumptions, e.g. entertainment-on-the-wheel [2]. The service-oriented vehicular networks have attracted a great deal of investment in both the development of in-vehicle devices and large-scale deployment of wireless infrastructures [3,4].

Today, wireless entertainment devices, e.g. mobile TVs, intelligence terminals, etc., are commonly equipped by vehicles. Service providers can publish popular multimedia contents through the RSUs to the OBUs in the *Area of Interest* (AoI), a process which is known as *popular content distribution* (PCD) [5,6]. The “popular contents” range from newly released movies, popular music, infotainments to

---

[*] Corresponding author at: School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China.
Tel.: +86 51188986871.
E-mail addresses: yzhw9981@ujs.edu.cn (W. Huang), wanglm@ujs.edu.cn (L. Wang).

http://dx.doi.org/10.1016/j.comnet.2016.02.006

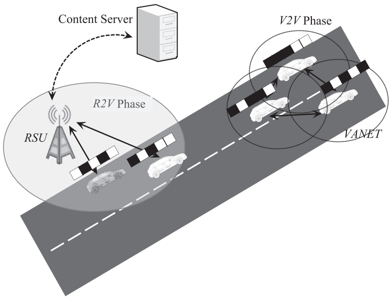


Fig. 1. Two phases of collaborative downloading.

local digital maps, and real-time traffic information, which are usually large files [7]. In the usual content downloading service, different OBUs download different files. However, in PCD, most of the OBUs in the AoI request the same content. Efficient downloading of popular content to drivers and passengers will be a critical marketing strategy for information service providers [8].

Existing solutions for PCD can be classified into two types: solutions based on cellular networks (GPRS, 3 G, 4 G, etc.) , and those based on DSRC [9,10]. The cellular network-based solutions have drawbacks in the high communication cost. Moreover, when the number of OBUs is large, the bandwidth is still limited. The DSRC based solutions suffer from the high cost of deploying a large number of hotspots at roadsides, providing the ubiquitous coverage envisaged with DSRC. In addition, communications between the vehicles and the hotspots may be frequently interrupted due to the high speed of the vehicles.

Collaborative downloading is a data dissemination method which allows the distribution of information among cooperative vehicles, and many researches [5–7,11,12] have indicated that collaborative PCD schemes can achieve lower downloading delay. In collaborative PCD, the content is divided into several generations (blocks) with equal size. As shown in Fig. 1, when the OBUs are within the coverage of the RSUs, they receive different generations, and this phase is known as the *R2V phase*; when the OBUs are out of the coverage of the RSUs, they form the VANET and exchange the generations, the phase known as the *V2V phase*. Those two phases alternate to complete the PCD.

Collaborative downloading on the Internet has been extensively studied, and several protocols (e.g. BitTorrent [13], eDonkey2000 [14], etc.) have been widely used and achieved strong performance. Due to the high speed of vehicles and the difficulty of predicting the vehicles' trajectories in urban scenarios, the direct application of the CD protocols of the application layer for the Internet to vehicular networks performs much worse. For example, for only 20 OBUs, the *SPAWN* protocol [15], a protocol similar to BitTorrent, takes almost 3 min to download a file only 1.6MB in size.

Rapid change and unpredictable topology are among the main characteristics of the VANETs, and this creates challenges for modeling urban vehicular networks [16]. Indeed, if treating the OBUs as nodes, the positions of the nodes are changing. Referring the idea of cellular division from the Cell Transmission Model (CTM) [17] and position-based routing strategies [18–20], each lane can be divided into successive homogeneous cells with the same size, and the OBUs in the same cell form a cluster. Therefore, the OBUs' movement on the road can be transformed to the action of leaving the old cell and joining the new cell. By treating the cells as nodes, the positions of the nodes remain unchanged, which greatly simplifies the modeling of the VANETs. IEEE 802.11p supports up to six service channels. Due to the recent researches and standardization (e.g. IEEE 1609.4) regarding dual ratio multichannel operation [21], intra-cluster generation dis-semination and inter-cluster generation exchange can be performed simultaneously, by using different ratios.

The relay selection and generation selection are two core issues that should be considered in a collaborative PCD scheme. Due to the limited number of channels, only a small number of the transmissions between the OBUs can be carried out simultaneously, otherwise severe packet collisions may be caused. Relay selection is the strategy to determine the priority of nodes, as only the nodes of high priority can access the channels and broadcast their generations. Due to the limited bandwidth, each node can only broadcast a small number of generations at each time step when accessing the channels. Generation selection is the strategy used to determine which generations one node should broadcast.

For the relay selection, the basic idea in our scheme ECDS is to make the high density and high downloading rate cells be of high probability to obtain the missing generations. As a result, the generations may be accumulated to some of the high density cells quickly, then be distributed by these cells to the surrounding cells. The basic idea under our scheme is not derived from experience alone, and the transmission process analysis from the complex network theory provides the theoretical support. The “high density first” principle ensures that the high density cells are of high degree, and the “high downloading rate first” principle ensures that the cells prefer to choose the cells of high degree as relays, thus leading to the *Matthew Effect* [22]. With these two principles, there are only a small number of cells of high degree, which we refer to as core cells. The majority of cells of low degree can connect to the core cells within a few hops, and this topology structure is similar to the *BA scale-free network* [23]. Pastor-Satorras and Vespignani [24] researched the epidemic spread process in different topologies, and discovered that the virus spread most quickly in scale-free networks. Recent researches [36,37] theoretically confirms their finding. PCD in VANETs is similar to the virus spread, and this is the chief reason ECDS is effective.

For the generation selection, the node accessing the channel first broadcasts the generation which the neighboring cells of the highest density and at high downloading rate require for, breaking the tie by broadcasting the generation which the most neighboring nodes require for.

Our contribution: In this paper, we address the V2V phase of collaborative PCD, and propose the ECDF, an efficient collaborate downloading scheme for popular content distribution in urban vehicular networks.

1. The cell-based clustering strategy is proposed. Each lane can be divided into successive homogeneous cells of equal size, and the OBUs in the same cell form a cluster. By treating the cells as nodes, the positions of the nodes are not changing, which greatly simplifies the modeling of the VANETs.
2. A topology pre-creation and update scheme is proposed, based on the cellular clustering. By introducing the cell density and degree into the preferential attachment probability, a topology of scale free property is built, which is beneficial for data dissemination.
3. A new inter-cluster relay selection strategy and generation selection strategy for the pre-created topology are proposed. By introducing the principles of "high cell density first" and "high downloading rate first", the generations are first accumulated to a small number of core cells, then distributed to a majority of low-degree cells within a few hops, which accelerates the PCD process.
4. Simulations in multiple scenarios are performed in this paper. By comparing the ECDS to two up-to-date collaborative PCD protocols in two recognized metrics, i.e. *average downloading percentage* [15] and *overall finish time* [25], we declare that the ECDS performs well in urban scenarios.

The remainder of the paper is organized as follows: In Section 2, we give a brief review of the recent researches on the content distribution schemes in VANETs, both for the V2V phase and R2V phase; in Section 3, the system model of ECDS is introduced; in Section 4, the simulation results of ECDS are given, comparing to two up-to-date collaborative PCD schemes; finally, in Section 5 we summarize ECDS and draw the conclusions of the study.

2. Related works

Research regarding the collaborative popular content distribution originates from Alok Nandan et al., who proposed the *SPAWN* protocol [15], in which a relay selection strategy known as "*Rarest-Closest First*" was proposed. One OBU first requests blocks from its nearest OBUs, breaking tie by selecting the OBUs possessing the rarest blocks. The drawback is that *SPAWN* uses TCP/IP as the underlying protocol and may cause high overhead. Ahmed and Kanhere first introduced network coding into a collaborative PCD scheme, and proposed *VANETCODE* [25]. In the R2V phase, the RSUs generate different coefficients for different OBUs, and send each OBU a single encoded block using liner coding, along with the coefficients. In the V2V phase, the OBUs exchange their encoded block and the coefficients, which are usually smaller than the original content and do not require generation selection. However, the number of the decoders required for decoding may be very large when downloading large contents, which may slow down the downloading process. Moreover, the computation of the decoding process is quite costly. Wu et al. [26] proposed a network coding and fuzzy logic based sender-oriented CD protocol, in which the relay selection strategy of their protocol is determined by the inter-vehicle distance, vehicle movement and signal strength; a simple retransmission method was used to deal with the interruption caused by the dynamic VANET topology. Li et al. [6] introduced the SLNC (*Symbol Level Network Coding*) into their collaborative PCD scheme, aiming at the lossy wireless links. They also proposed a new relay selection strategy based on "*node utility*", but the analytical explanation regarding why their strategy is effective is not given. Liu and Chen [27] researched a special case of the collaborative PCD. They assumed that some vehicles park at the roadsides, and these stopped vehicles can be used as relays. Concretely, the stopped vehicles can be treated as a strip-like cluster, thus the communication between the moving vehicles and the parked vehicles can be abstracted as communications between the moving points and the static line. Their research provides us with inspiration for treating the cluster as nodes. More recently, Wang et al. [7] proposed a coalition formation game based solution for collaborative PCD. For every K time slot, the network is divided into several sub-networks. For each sub-network, the relay selection is determined by the Nash equilibrium of the coalition formation game, and the generation selection is based on a proposed greedy strategy. However, the performance of their protocol becomes worse when the content size is large.

Many works focus on the R2V phase of the collaborative PCD, all of which assume a perfect V2V content exchange. Panwai and Dai [1] further studied the R2V phase in VANETCODE, and proposed the feedback-based mechanism. They also performed an analytical derivation in the case of two vehicles. Trullols- Cruces et al. [16] focus their work on the R2V phase, using the vehicles' position prediction, which is based on the trajectory mining, to create the "contact map" in the future. With the help of the predicted contact map, the selection of the carrier and scheduling of the data chunks strategies can be effective. The drawback is that their methods are limited by the accuracy of position prediction, which is difficult to guarantee. Zhu et al. [11] proposed a feedback-based scheme for the R2V phase in collaborative downloading, and network coding is also used in their protocol. For additional state-of-the-art technologies and protocols for content distribution in VANETs, refer to Gerla's survey [28].

Only a small number of works introduce complex network theories to the VANET topology control. The chief reason for this is that it is a challenge to build a model of the rapid changing topology of VANETs on the basis of evolution models. Schleich et al. [29] proposed a method to build the VANET topology with the concept of small world networks, and used the NSGAII algorithm to minimize the APL (*Average Path Length*) and maximize the CC (*Clustering Coefficient*). Their method cannot guarantee that the topology is of small world property at each time step. Moreover, their method does not perform well when the speeds of the vehicles are high. Banerjee et al. [30] proposed a method to build the ad-hoc network with small world property, in which "*short-cuts*" are created by using directional antennas stems, but the mobility of the nodes is not considered in their method. None of those methods are designed for collaborative downloading, which renders them unsuitable for PCD in urban VANETs.

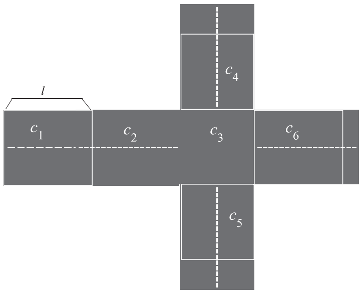


Fig. 2. Cell-based cluster division.

3. System model

In this section, we first introduce the cell-based clustering scheme; then we introduce the intra-cluster generation distribution process and inter-cluster generation exchange process, with the analytical derivation on the degree distribution of the topology and a differential equation modeling of the generations exchange process.

3.1. Overview: the ECDS scheme

We generally describe the proposed collaborative PCD scheme for urban VANETs as follows: each vehicle is equipped with an OBU including two wireless transceivers (dual radios) operating on multiple channels. With the cell-based clustering scheme, the OBUs join in different clusters depending on their geographical positions. At every T time step, the topology pre-creation (periodical update) shall be carried out. At each time step, the OBUs in each cluster compete for the intra-cluster channel according to the utility value; when one OBU accesses the channel, it broadcasts the generation selected by the intra-cluster generation selection strategy. The CH of each cluster competes for the inter-cluster channels according to the cell priority, and when it accesses the channel, it broadcasts the generation determined by the inter-cluster generation selection strategy. Utilizing multi-interface multi-channel technology [21], these two processes can be performed simultaneously to decrease the delay.

3.2. Cell-based clustering scheme

As shown in Fig. 2, each lane is divided into continuous homogeneous cells with equal length l, and the crossings are treated as independent cells. The OBUs in the same cell form a cluster. Assuming that the vehicles are equipped with GPS devices, this operation is easy to realize.

There are three advantages to treating the cells as nodes. First, the vehicles in the same cell are located quite close to each other in the same lane. Referring to the researches of the car-following model [1], these vehicles usually have similar movement patterns. As a result, the connections between the vehicles in the same cluster may be interrupted less frequently. Second, based on cellular division, the cell density can be defined and used to estimate the density of some area, which facilitates the relay and generation section. Third, rather than treating the OBUs as nodes, treating the cells as nodes can eliminate the changing of the nodes' positions. The vehicles' moving changes to the action of leaving the old cluster and joining the new one, but the positions of the clusters remain unchanged.

The density of cell c in a short time interval is defined as the average number of OBUs that stay on cell c, signed as $\rho(c)$. Each cluster elects a CH (cluster head), and it is reasonable to select the vehicle of the lowest speed or which will stay in the cell for the longest time to be the cluster head [31]. CHs are responsible for calculating the cell density, organizing the distribution of intra-cluster generations, and coordinating the inter-cluster generation exchange. In addition, as cell density cannot be suddenly changed in urban scenarios, we can build the topology in advance with the inter-cluster relay selection strategy, and update it periodically. In this manner, redundant wireless links are erased, thereby reducing communication overhead.

3.3. Intra-cluster generation distribution

Assuming the file F is divided into L generations G_1, G_2, … , G_L, the size of each generation is M, and the bandwidth is B. For each time step, only $K = B/M$ generations can be broadcast in each cluster. Similar to the research of Li [6], the intra-cluster relay selection and generation selection is based on the *utility value*.

We let Ω_i be the set of generations possessed by i, and for generation G_j, the utility of i, $U(i, G_j)$ is defined as the number of OBUs in the cluster which do not possess the generation G_j, formulized as follows:

$$U(i, G_j) = |\{k|G_j \notin \Omega_k\}| \quad (1)$$

The utility of i, $U(i)$, is defined as the maximum value of $U(i, G_j)$, $1 \le j \le L$. The generation i should be broadcast when it enters the channel, g_i^*, and is the generation with the maximum value of $U(i, G_j)$, $1 \le j \le L$.

$$U(i) = \max_{1 \le j \le L} U(i, G_j) \quad (2)$$

$$g_i^* = \arg\max_{1 \le j \le L} U(i, G_j) \quad (3)$$

In each time step, the OBUs with the highest K $U(i)$ values can access the channel and broadcast the generation g_i^*.

The intra-cluster generation distribution process is simple. In case of collision, the neighboring clusters should use different data channels. The newly joined OBUs should broadcast the vector, indicating which generations it possesses. At each time step, each OBU in the same cluster calculates its utility and continues broadcasting its utility through the control channel. When the data channel is not occupied, the OBU with the highest utility accesses the data channel and broadcasts its generation determined by Eq. (3).

3.4. Inter-cluster generation exchange

3.4.1. Two principles

The schemes for the inter-cluster generation exchange are based on the two principles of high downloading rate first and high cell density first.

The high cell density first principle means that the high cell density clusters are of high probability to obtain the missing generations. Two reasons may support the high cell density first principle. First, recent experimental research (Sumalee et al. [33], Zhang et al. [34], Moylan et al. [35], etc.) on macroscopic traffic flow have shown that actual traffic patterns obey the triangular shaped flow-density diagram, which indicates that the flow becomes slow when a cell's density becomes high. In other words, the speeds of the vehicles in high density cells are slow. Therefore, the vehicles may remain in the cell for a comparatively longer time, which benefits the generations' accumulation and distribution. Second, the number of vehicles in the high density cells is greater than those in the other cells. Therefore, accumulating generations to the cells of high density may benefit a majority of vehicles acquiring the generations.

The high downloading rate first principle means that the high downloading rate clusters (or the clusters which possess more generations) are of high probability to acquire the missing generations. This principle leads to the *Matthew Effect*: the nodes which possess more generations than the others can accumulate generations faster and faster. From the perspective of topology, if node (or cell) u exchange blocks with node v between two topology update operations, then an edge exists between u and v. One node of high downloading rate has a high degree of topology, and it also has a faster degree increasing rate than others. As a result, a small number of nodes are of high degrees, while the majority of the nodes are of low degrees, but can touch the high degree nodes in a few hops, which is the typical characteristic of the scale-free network. Information spreads faster in a scale-free network [24,36,37], which is why ECDS is effective.

3.4.2. Topology pre-creation and periodical update

By treating the cells as nodes, the positions of the nodes can be fixed, and the cell's density can be treated as a constant in a short time interval. Due to the two proposed principles, some pairs of nodes can only exchange generations at a very low probability in a certain time period. Therefore, the topology can be explicitly created in advance to reduce unnecessary wireless links. As the cells' densities may change significantly after a long period of time, the topology should be updated periodically.

Concretely, similar to the evolution process of a BA scale-free network, the creation process begins from a particular node (the root), e.g. the node containing RSUs. Initially, the aggregate only contains the root and its neighboring nodes. At each step, as shown in Fig. 3, one node adjacent to the nodes in the aggregation joins in and establishes m links to the nodes existing in the aggregate, according to the preferential attachment probability; it is suggested that $m = 3$. VANETs are not the relational networks (e.g. Internet or social networks) due to the limited communication distance. In relational networks, one node can establish links to any other nodes. However, in VANET, the newly joined node can only establish links with the adjacent ones. This process ends when all of the nodes are in the aggregate. Then the topology is created.

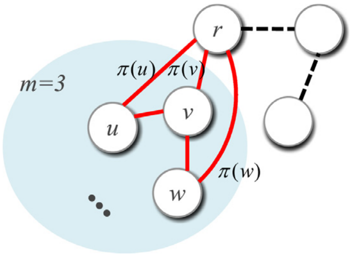


Fig. 3. Illustration of the topology pre-creation process, in which the blue nodes are the nodes in the aggregation, and the nodes connected by dotted lines are the nodes which are adjacent. At step s, r joins in the aggregation and establishes one edge to u, v and w respectively, with probability $\pi(u)$, $\pi(v)$, $\pi(w)$.

The preferential attachment probability is proportional to the product of the node's degree and the density, formalized as follows:

$$\pi(v_i) = \frac{d(v_i)\rho(v_i)}{\sum_{v_j \in adj(v_k)} d(v_j)\rho(v_j)} \tag{4}$$

Where $d(v_i)$ is the degree of node v_i.

When one cell is joining in the aggregation, it requests the densities and degrees of the cells both adjacent to it and in the aggregate. After calculating each cell's attachment probability, it uses roulette wheel selection [32] m times to establish m links.

By introducing the product of one node's degree and density into the preferential attachment probability, the nodes of high density in the aggregate are more likely to establish links with the newly joined node, thus the degree of some high density node may be greater than the other nodes. As the degree is a factor of the preferential attachment probability, the high density nodes of higher degree have higher probability to establish links, i.e. to acquire the generations from the other nodes. From the above analysis, this process directly reflects the principles of high downloading rate first and high cell density first. Furthermore, we prove that the topology created by the process is of scale-free property.

After a certain amount of time, the densities of some cells may be quite different, at which point the node containing RSUs should restart the topology creation process, which is known as the topology periodical update.

3.4.3. Degree distribution derivation and scale-free validation

In this section, the analytical derivation of the degree distribution is given in order to prove the scale-free property. The mean field differential equation on the degree is as follows:

$$\frac{\partial d(v_i)}{\partial s} = m \cdot P(v_i \in adj(v_j)) \cdot \frac{d(v_i)\rho(v_i)}{|adj(v_j)| < d(s) > \overline{\rho}} \tag{5}$$

Where $adj(v_j)$ is the set of adjacent nodes of v_j, $<d(s)>$ is the average degree of the nodes in the aggregation at step s, v_j is the new joined node at step s, and $\bar{\rho}$ is the average density of the cells.

As the layout of the roads is regular, the number of adjacent nodes of one node is usually as same. Therefore, we can approximate $|adj(v_j)|$ to a constant A, after which we

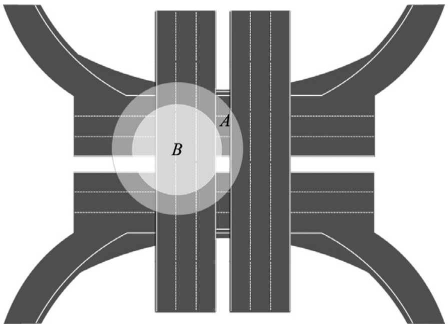


Fig. 4. Limited degree increase.

have:

$$P(v_i \in adj(v_j)) = \frac{A}{s} \tag{6}$$

$$\langle d(s) \rangle = \frac{1}{s} \sum_{v \in V^{(s)}} d(v) \tag{7}$$

$$\sum_{v \in V^{(s)}} d(v) = 2ms \tag{8}$$

Where $V^{(s)}$ is the vertex set of the aggregation at step s. Substituting (8) into (7), we have:

$$\langle d(s) \rangle = 2m \tag{9}$$

Then, substituting (6), (7), (8) and (9) into (5), we have:

$$\frac{\partial d(v_i)}{\partial s} = \frac{d(v_i)\rho(v_i)}{2\overline{\rho} s} \tag{10}$$

According to the initial condition $d(v) = m$ when $s = s_i$, by solving the differential equation, we have the following:

$$d(v_i) = m\left(\frac{s}{s_i}\right)^{\beta(v_i)} \tag{11}$$

Where $\beta(v_i) = \rho(v_i)/2\bar{\rho}$.

It can be seen that the degrees of the nodes in the aggregation increase exponentially at different rates. Differing from the evolution of the BA scale-free network, the nodes which join V at the early steps may stop increasing in degree. Fig. 4 is an example in which the outer circle is the aggregation at some same time step, and the newly joined nodes in the future steps can only communicate with the nodes in area A. In other words, the nodes in area B will stop growth in degree.

For any node v_i, a stopping step u_i exists, namely:

$$d(v_i) = m\left(\frac{s}{s_i}\right)^{\beta(v_i)} \le m\left(\frac{u_i}{s_i}\right)^{\beta(v_i)} \tag{12}$$

With the increase of s, every node in V will reach the degree upper limit. Therefore, we have the following formulas:

$$d(v_i) = m\left(\frac{u_i}{s_i}\right)^{\beta(v_i)} \approx m\left(1 + \frac{\overline{u}}{s_i}\right)^{\beta(v_i)} \tag{13}$$

Where

$$\overline{u} = \frac{1}{|V_s|} \sum_{v_i \in V_s} (u_i - s_i) \tag{14}$$

Let $d(v_i) < k$, and according to (11), we have:

$$m\left(1 + \frac{\overline{u}}{s_i}\right)^{\beta(v_i)} < k \tag{15}$$

For $k \ge m$, (15) is equivalent to:

$$s_i > \overline{u} \Bigg/ \left[\left(\frac{k}{m}\right)^{\frac{1}{\beta(v_i)}} - 1\right] \tag{16}$$

By introducing a small constant ε, (16) can be approximated to:

$$s_i > \overline{u} \cdot \left(\frac{k}{m(1+\varepsilon)}\right)^{-\frac{1}{\beta(v_i)}} \tag{17}$$

Then we have:

$$P(d(v_i) < k) = P\left(s_i > \overline{u} \cdot \left(\frac{k}{m(1+\varepsilon)}\right)^{-\frac{1}{\beta(v_i)}}\right) \tag{18}$$

As v_i can join the aggregation at any step from [1, s], thus s_i obeys uniform distribution $U(1, s)$, $p_s(s_i) = 1/s$:

$$P(d(v_i) < k) = 1 - \frac{\overline{u}}{s}\left(\frac{k}{m(1+\varepsilon)}\right)^{\frac{-1}{\beta(v_i)}} \tag{19}$$

$$P(d(v_i) = k) = \frac{\partial P(d(v_i) < k)}{\partial k} = \frac{1}{\beta(v_i) m(1+\varepsilon)} \cdot \frac{\overline{u}}{s} \cdot \left(\frac{k}{m(1+\varepsilon)}\right)^{-\frac{1}{\beta(v_i)} - 1} \tag{20}$$

Substituting $\beta(v_i) = \rho(v_i)/2\bar{\rho}$ into (20), we have:

$$P(d(v) = k | \rho(v) = \rho_0) = \frac{2\overline{\rho}}{\rho_0 m(1+\varepsilon)} \cdot \frac{\overline{u}}{s} \cdot \left(\frac{k}{m(1+\varepsilon)}\right)^{-\frac{2\overline{\rho}}{\rho_0} - 1} \tag{21}$$

Let $p_\rho(\rho_0)$ be the distribution of cell density. According to the formula of full probability, we can derive the degree distribution function $p(k)$:

$$p(k) = \int_0^{\max \rho} p_\rho(\rho_0) \cdot p(d(v) = k | \rho(v) = \rho_0) d\rho_0 = \int_0^{\max \rho} p_\rho(\rho_0) \cdot \frac{2\overline{\rho}}{\rho_0 m(1+\varepsilon)} \cdot \frac{\overline{u}}{s} \cdot \left(\frac{k}{m(1+\varepsilon)}\right)^{-\frac{2\overline{\rho}}{\rho_0} - 1} d\rho_0 \tag{22}$$

According to the integral mean value theorem, ρ' ranging from [0, max p] exists, and we have the following equation:

$$\int_0^{\max \rho} p_\rho(\rho_0) \cdot \frac{2\overline{\rho}}{\rho_0 m(1+\varepsilon)} \cdot \frac{\overline{u}}{s} \cdot \left(\frac{k}{m(1+\varepsilon)}\right)^{-\frac{2\overline{\rho}}{\rho_0} - 1} d\rho_0 = \max \rho \cdot p_\rho(\rho') \cdot \frac{2\overline{\rho}}{\rho' m(1+\varepsilon)} \cdot \frac{\overline{u}}{s} \cdot \left(\frac{k}{m(1+\varepsilon)}\right)^{-\frac{2\overline{\rho}}{\rho'} - 1} \tag{23}$$

Eqs. (22) and (23) clearly indicate that the degree distribution of the created topology obeys power-law distribution, thereby proving that the topology created is of scale-free property.

3.4.4. Inter-cluster relay selection & generation selection

Let $G(V, E)$ be the aggregation, where V is the set of nodes (clusters) in the aggregation and E is the set of edges.

For the relay selection, following the principles of high downloading rate first and high cell density first, the priority of node v, $pri(v)$, is defined as the maximum value of the products of $d(u)$ and $\rho(u)$, where u is the neighboring node of v, and v possesses the generations u does not. Breaking tie by selecting the node with highest utility, we formulize the following:

$$pri(v) = \max_{u \in adj(v), \exists G_i G_i \in \Omega_v \wedge G_i \notin \Omega_u} d(u)\rho(u) \tag{24}$$

For the generation selection, one node should broadcast the generation which the neighboring node with the highest $d(u)\rho(u)$ value does not possess, breaking the tie by broadcasting the generation which the most of the neighboring nodes require, formulized as follows:

$$g_v^* = \mathop{\arg\max}_{u \in adj(v), \exists G_i G_i \in \Omega_v \wedge G_i \notin \Omega_u} d(u)\rho(u) \tag{25}$$

By assigning the high propriety to the nodes whose neighboring nodes are of high $d(u)\rho(u)$ value, these nodes have higher priority to accessing the channel, and first broadcast the generation which the neighboring nodes with the highest $d(u)\rho(u)$ value request. Therefore, the nodes of high $d(u)\rho(u)$ value are of high probability to acquire the missing generations. Based on the above analysis, it is shown that the inter-cluster relay selection and generation selection strategies follow the principles of high downloading rate first and high cell density first.

The inter-cluster generation exchange process is similar to the intra-cluster generation distribution process. At each time step, the CH of each cluster broadcasts its priority and the vector indicating which generations it possesses through the control channel. When one of the inter-cluster data channels is not occupied, then the node of the highest priority enters the channel and broadcasts its generations. The nodes adjacent to it in the aggregation turn to the same channel and receive the missing generations.

3.4.5. Analysis of generation exchange process

In this section, we analyze the generation exchange process in the time interval between two topology update operations. As this time interval is very short (less than 10 s), it is reasonable to assume that the CH of each cluster does not change. Moreover, the impact of intra-cluster distribution can be ignored. Without the loss of generality, we assume that $B = M$, which means that a cluster can acquire at most one missing generation at each time step; for the case $B > M$, each time step can be simply divided into B/M time slices, and the following analysis still applies.

One cluster can acquire one missing generation if and only if:

(1) The cluster has the highest product of density and degree among the neighboring clusters.
(2) The neighboring clusters possess the generation the cluster does not.

Let $GP\ (n,\ d',\ \rho',\ \tau)$ be the probability that one cluster possesses n generations, at time τ whose degree is d' and cell density is ρ'. At time step τ, for a node v possessing n' generations, the probability that the neighboring clusters possess any generation that v does not is:

$$\begin{aligned} &P_O(n', d(v), \tau) = P(\exists u \in adj(v), \Omega_u \not\subset \Omega_v) \\ &= 1 - P(\forall u \in adj(v), \Omega_u \subset \Omega_v) \\ &= 1 - \left\{ \sum\nolimits_j \left\{ p_1(j|d(v)) \sum\nolimits_{\rho'} \left[\begin{array}{c} p_2(\rho'|j) \cdot \\ \sum\nolimits_n \left(GP(n, j, \rho', \tau) \cdot \frac{C^n_{n'}}{C^n_L} \right) \end{array} \right] \right\} \right\}^{d(v)} \end{aligned} \tag{26}$$

The probability that the cluster v has the highest priority among the competing neighboring clusters is as follows. The competing clusters of v are the clusters which are both adjacent to v and are able to acquire the missing generations from their neighboring nodes.

$$\begin{aligned} &P_H(\rho(v), d(v), \tau) = P\left(\rho(v) \cdot d(v) \geq \max_{\substack{u \in adj(v) \\ \exists w \in adj(u), \Omega_w \not\subset \Omega_u}} \rho(u) \cdot d(u) \right) \\ &= P\left(\forall u \in adj(v) \left(\begin{array}{c} \rho(u) \cdot d(u) \leq \rho(v) \cdot d(v) \vee \\ \forall w \in adj(u), \Omega_w \subset \Omega_u \end{array} \right) \right) \\ &= \left[\begin{array}{c} \sum\nolimits_{i \cdot \rho' \leq \rho(v) \cdot d(v)} p_1(i|d(v)) \cdot p_2(\rho'|i) \\ + \sum\nolimits_{i \cdot \rho' > \rho(v) \cdot d(v)} \left(\begin{array}{c} p_1(i|d(v)) \cdot p_2(\rho'|i) \cdot \\ \sum\nolimits_n GP(n, i, \rho', \tau)(1 - P_O(n, i, \tau)) \end{array} \right) \end{array} \right]^{d(v)} \end{aligned} \tag{27}$$

For any node v whose degree is k_2, when randomly selecting one node u adjacent to v, the probability that the degree of u is k_1 is signed as $p_1(k_1|k_2)$. Referring to the branching process in complex network [38], we have:

$$p_1(k_1|k_2) = \frac{k_1 p(k_1)}{\langle k \rangle} \tag{28}$$

Which is independent of k_2. and $p_2(\rho'|i)$ is the probability that one node of degree i has the density ρ'. According to (22), we have:

$$\begin{aligned} p_2(\rho'|i) &= p_\rho(\rho') \cdot \frac{2\overline{\rho}}{(1+\varepsilon)\rho' m} \cdot \frac{\overline{u}}{s} \\ &\cdot \left(\frac{i}{m(1+\varepsilon)} \right)^{-\frac{2\overline{\rho}}{\rho'} - 1} \bigg/ p(i) \end{aligned} \tag{29}$$

Now we can obtain the difference equation for the GP.

$$\begin{aligned} &GP(n, d', \rho', \tau + 1) \\ &= GP(n-1, d', \rho', \tau) P_H(\rho', d', \tau) P_O(n-1, d', \tau) \\ &+ GP(n, d', \rho', \tau)\left[1 - P_H(\rho', d', \tau) P_O(n, d', \tau)\right] \end{aligned} \tag{30}$$

At time step τ, the expectation of the generation quantity of the node whose degree is d' and cell density is ρ' is as follows:

$$GE(d', \rho', \tau) = \sum\nolimits_n n \cdot GP(n, d', \rho', \tau) \tag{31}$$

The expectation of the generation quantity of a node at time step τ can be obtained with the integral on d' and ρ':

$$GE(\tau) = \iint_{d', \rho'} p(d') p_2(\rho'|d') GE(d', \rho', \tau) \tag{32}$$

We next attempt to prove the property that the node with higher cell density has the higher probability to acquire the generations at each time step.

Property 1. The node with higher cell density has higher probability to acquire the generations from the neighboring nodes.

For one node whose degree is d' and cell density is ρ', and possesses the n' generations, the probability that the node can acquire one more generation from its neighboring nodes at time step τ is:

$$PG(n', d', \rho', \tau) = P_O(n', d', \tau) \cdot P_H(\rho', d', \tau) \tag{33}$$

We must prove that $PG(n', d', \rho_1, \tau) > PG(n', d', \rho_2, \tau)$, where $\rho_1 > \rho_2$. As P_O is independent from the cell density, we only need to prove that $P_H(\rho_1, d') > P_H(\rho_2, d')$:

$$\begin{aligned} &P_H\left(\rho_1, d'\right)/P_H\left(\rho_2, d'\right) \\ &= \left[\frac{\begin{array}{c}\sum_{i\cdot\rho' \le \rho_1 \cdot d'} p_1(i|d') \cdot p_2(\rho'|i) + \\ \sum_{i\cdot\rho' > \rho_1\cdot d'} \left(\begin{array}{c} p_1(i|d') \cdot p_2(\rho'|i) \cdot \\ \sum_n GP(n, i, \rho', \tau)(1 - P_O(n, i, \tau))\end{array}\right)\end{array}}{\begin{array}{c}\sum_{i\cdot\rho' \le \rho_2 \cdot d'} p_1(i|d') \cdot p_2(\rho'|i) + \\ \sum_{i\cdot\rho' > \rho_2\cdot d'} \left(\begin{array}{c} p_1(i|d') \cdot p_2(\rho'|i) \cdot \\ \sum_n GP(n, i, \rho', \tau)(1 - P_O(n, i, \tau))\end{array}\right)\end{array}}\right]^{d'} \\ &= \left[1 + \frac{\sum_{i\cdot\rho' \in (\rho_1\cdot d', \rho_2 \cdot d')} \left[\begin{array}{c} p_1(i|d') \cdot p_2(\rho'|i) \cdot \\ (1 - \sum_n GP(n, i, \rho', \tau)(1 - P_O(n, i, \tau)))\end{array}\right]}{\begin{array}{c}\sum_{i\cdot\rho' \le \rho_2 \cdot d'} p_1(i|d') \cdot p_2(\rho'|i) \\ + \sum_{i\cdot\rho' > \rho_2\cdot d'} \left(\begin{array}{c} p_1(i|d') \cdot p_2(\rho'|i) \cdot \\ \sum_n GP(n, i, \rho', \tau)(1 - P_O(n, i, \tau))\end{array}\right)\end{array}}\right]^{d'} \\ &> 1 \end{aligned} \tag{34}$$

Therefore, we have $P_H(\rho_1, d') > P_H(\rho_2, d')$, and Property 1 is proven.

Next, we attempt to prove the property that the node with higher degree has the higher probability to acquire the generations at each time step.

Property 2. The node with higher degree has higher probability to acquire the generations from the neighboring nodes.

We must prove that $PG(n', d_1, \rho', \tau) > PG(n', d_2, \rho', \tau)$, where $d_1 > d_2$. As $p_1(i|d')$ is independent from d', similar to the proof in Property 1, we can obtain $P_H(\rho', d_1, \tau) > P_H(\rho', d_2, \tau)$ and $P_O(n', d_1, \tau) > P_O(n', d_2, \tau)$.

According to the above properties, we can see that the inter-cluster generation exchange processes do in fact follow the principles of "high density first" and "high downloading rate first" which we proposed.

3.5. ECDS protocol

As has been introduced above, the ECDS protocol consists of three major processes, based on the cellular clustering scheme: topology pre-creation, intra-cluster generation distribution and inter-cluster generation exchange. The pseudo-codes of these processes are listed as follows.

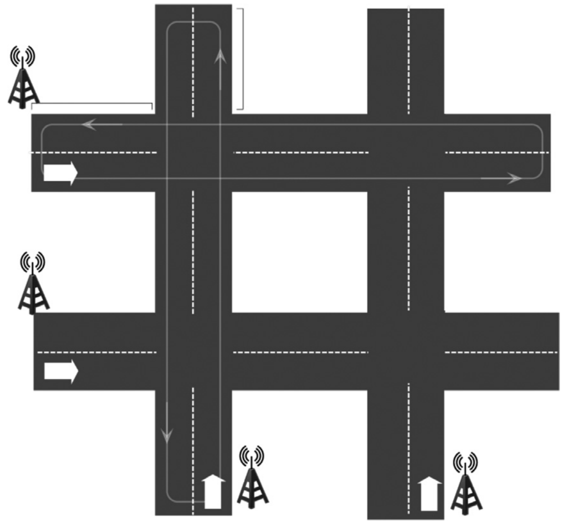


Fig. 5. The road network.

4. Simulation

In this section, the performance of the proposed scheme ECDS is evaluated and compared to two up-to-date protocols: the first is CFG [7], a dynamic popular content distribution protocol using coalition formation games for freeway VANETs; and the second is CodeOn [6], a cooperative popular content distribution protocol for VANETs.

As multi-interface multi-channel technology is used in ECDS, which has not been supported by most of the popular network simulators (e.g. NS2 and OPNET). As a result, we have developed a MATLAB-based discrete time simulator, and programmed these three protocols with MATLAB. For the sake of evenness, the multi-interface multi-channel technology is applied to all of the comparison protocols. Specifically, the Symbol Level Network Coding (SLNC) method used by CodeOn is not covered in the experiences. As SLNC is a method for error correction, which is irrelevant in our subject – this paper focus on how efficient relay selection and generation selection strategy can speed up the downloading process, rather than on the error correction technology itself. For lossy channels, we can simply apply the SLNC to any of the protocol.

4.1. Simulation setup

As shown in Fig. 5, we take the grid-like road network for the experiment, in which the length of each road segment is 600 m, and the RSUs are located at four points. Initially the network is empty, and vehicles enter the network from the points *A*, *B*, *C* and *D*. The vehicles do not exit the network when they reach the terminal of the road, instead they turn around and go in the opposite direction. We use the professional microscope traffic simulator VISSIM 5.20 to generate the trajectories in order to simulate the realistic traffic flow's movement. For the R2V phase, the RSUs randomly broadcast two generations at each time step (see Table 1).

Table 1
The ECDS protocol.

The ECDS protocol

```
  for every $T$ time steps
    call The topology pre-creation process
end
for each time step $t$
  for the CH of each cell
    call The inter-cluster generation distribution
  end
  for each OBU
    call The intra-cluster generation distribution
  end
End
```

The topology pre-creation process

```
The cells $v$ containing RSUs:
  $aggregation = \{v\}$
  randomly select one node $u$, $u \in adj(v)$
  send $(aggregation, u)$
The cells $v$ not containing RSUs:
  while $aggregation =$ receive ()
    for $u_i$ in $aggregation$
      query $(d(u_i), \rho(u_i))$
  end
    Select $m$ nodes $w_1, w_2, \ldots, w_m$ using roulette wheels selection according to the preferential attachment probability in (4).
    create_links $(v, w_i)$, $1 \le i \le m$
    randomly select one node $w$, $w \in adj(v)$ and $w \notin aggregation$
    $aggregation \leftarrow aggregation \cup \{v\}$
  send $(aggregation, w)$
End
```

The Intra-cluster generation distribution

```
  do
    $list = \{u_1, u_2, \ldots, u_k\}$, where $u_i$ and $v$ are in the same cell, $u_i \ne v$
    for $u_i$ in list
      query $(\Omega(u_i))$
    end
    calculate $U(v)$ according to (2)
    for $u_i$ in $list$
      query $(U(u_i))$
    end
  if $U(v) > \max\{U(u_i)\}$, $1 \le i \le k$ and intra-cluster channel is available
    broadcast ($g_v^*$, intra-cluster interface), where $g_v^*$ is calculated according to (3)
  else
    receive (intra-cluster interface)
    update $\Omega_v$
    end if

  while $\tau \in (t, t+1)$
```

The Inter-cluster generation distribution

```
  for every $T$ time steps
    $list = \{u_1, u_2, \ldots, u_k\}$, where $v$ has link to $u_i$, $1 \le i \le k$
    for $u_i$ in $list$
      query $(d(u_i)\rho(u_i))$
    end
  end
  Do
    for $u_i$ in $list$
      query $(U(u_i))$
    end
    calculate $pri(v)$ according to (24)
    for $u_i$ in $list$
      query $(pri(u_i))$
    end
    if $pri(v) > \max\{pri(u_i)\}$, $1 \le i \le k$ and inter-cluster channel is available
      broadcast ($g_v^*$, inter-cluster interface), where $g_v^*$ is calculated according to (25)
    else
      receive (inter-cluster interface)
      update $\Omega_v$
  end
while $\tau \in (t, t + 1)$
```

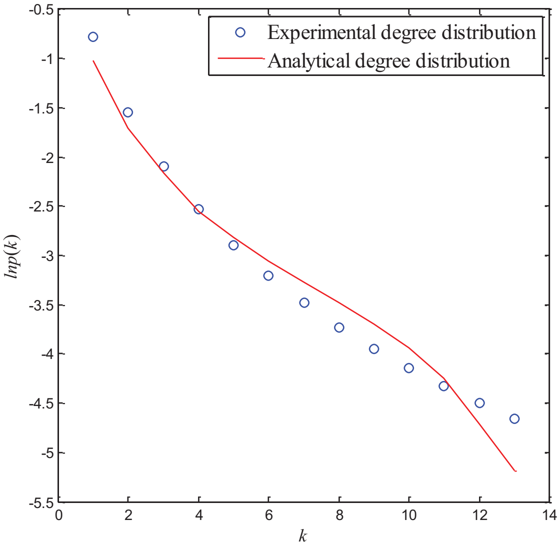


Fig. 6. Experimental vs. analytical degree distribution.

The experiments are carried out in multiple scenarios of different vehicles′ speeds and different content sizes. The detailed parameters are listed in Table 2.

4.2. Simulation 1 – generation exchange process

First, the inter-cluster generation exchange in ECDS is simulated separately in a small time interval of 5 s, between one topology periodical update and the next. We compare the simulation results to the analytical results in the degree distribution and the expected generation quantity.

Fig. 6 shows the degree distribution, the line shows the analytical distribution, and the discrete points show the experimental degree frequency. It can be seen that the experimental distribution meets the analytical distribution well. Moreover, the approximate linear relationship between k and $\ln(p(k))$ clearly indicates that the degree distribution obeys power law distribution.

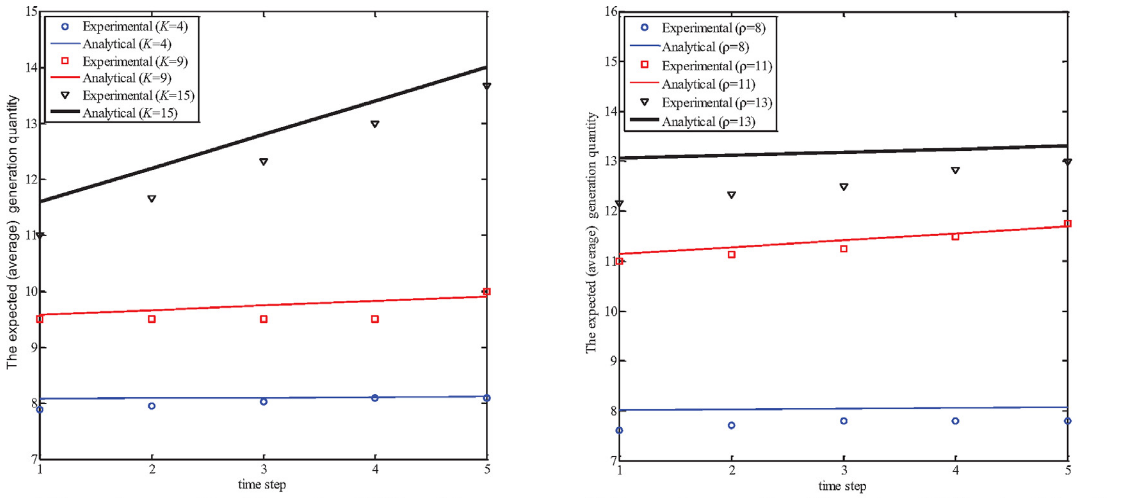
Table 2
The parameters of ECDS.

Parameters	Value
Cell size	150 m
Time step	1 s
The coverage radius of the OBUs	600 m
Bandwidth	16 Mbit/s
Generation size	8Mbit
File sizes	64, 128, 256MB
Speeds	24, 40, 60 km/h
Traffic flow rate	3600 veh/h
Acceleration	1 m/s^2
The probability of changing speed	0.1
Security distance	2 m
The periodicity of topology update	10 time steps
Simulation duration	600 s

Fig. 7(a) shows the analytical expected generation quantity and the experimental average generation quantity of the nodes of different degrees. The analytical expected generation quantity can be calculated from (30), and the experimental average generation quantity can be obtained from the statistics of the simulation results. It can be seen that the nodes with higher degree acquire more generations at each time step. This result strongly demonstrates that the proposed scheme reflects the principle of “high downloading rate first”.

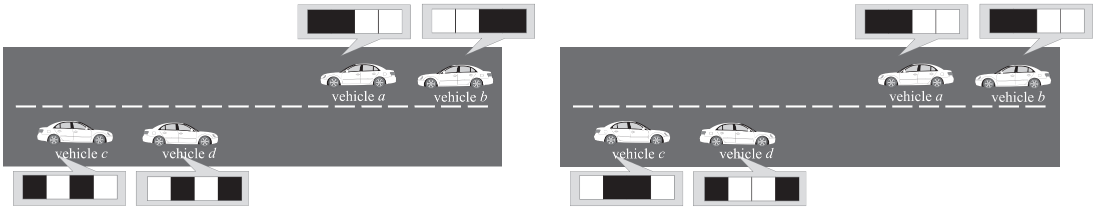
Fig. 7(b) shows the analytical expected generation quantity and the experimental average generation quantity of the nodes of different densities. It can be seen that the clusters of higher cell density possess more generations at each time step, which demonstrates that the proposed scheme reflects the principle of “high density first”.

Indeed, there are some deviations between the analytical expected generation quantity and the experimental average generation quantity. In most cases, the experimental average generation quantity is smaller than the analytical. The reason for this is that in the analytical reduction, only



a) Generation quantity of the nodes of different degrees b) Generation quantity of the nodes of different cell densities

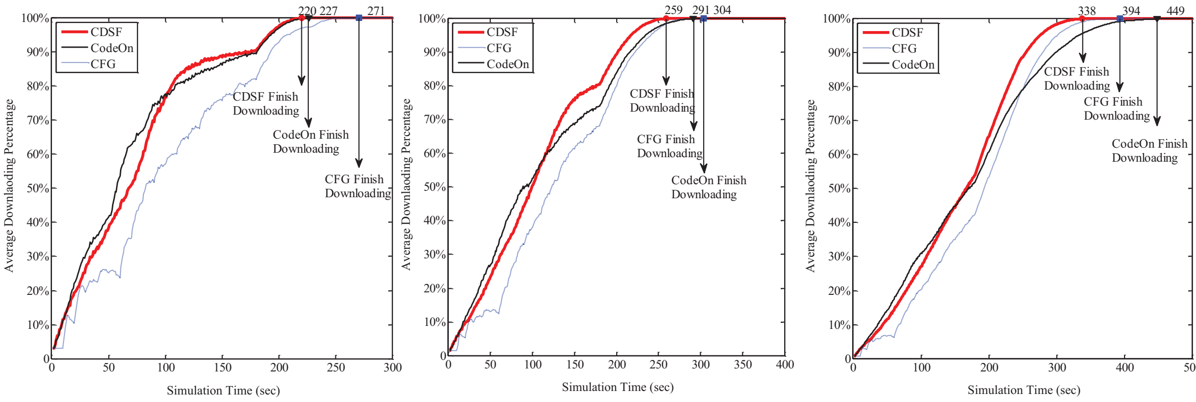
Fig. 7. Analytical expected generation quantity vs. experimental average generation quantity.



a) Uniform generations distribution

b) Non-uniform generations distribution

Fig. 8. Uniform and non-uniform generations distribution.



a) Content size = 64MB

b) Content size = 128MB

c) Content size = 256MB

Fig. 9. The Average Downloading Percentage (*ADP*) in downloading the contents of different sizes.

the numbers of the generations are considered, and the assumption is made that the generations are distributed uniformly among the nodes. However, in the simulation, the assumption may not always be established. The non-uniform distribution does in fact slow down the generation distribution process.

Let us look at a detailed example; assuming the content is divided into four generations, Fig. 8(a) shows the uniform distribution of the generations (the appearance times of different generations are equal), and Fig. 8(b) shows the non-uniform distribution (it can be seen that the second generation appears three times but the final generation only once). In Fig. 8(a), it is shown that vehicles *a* and *b* make a generation exchange, as do vehicles *c* and *d*, then all of the vehicles are able to obtain the entire content. However, in Fig. 8(b), the generation exchange process is more complex and requires more time than in the previous case.

4.3. Simulation 2 – performance evaluation

In this section, we simulate the three schemes: ECDS, CodeOn and CFG in the time interval [0, 600]. Two recognized metrics, i.e. *Average Downloading Percentage* (*ADP*) [15] and *Overall Finish Time* (*OFT*) [25], are used to evaluate the performance of the protocols. *ADP* is the mean of the downloading percentage of all of the OBUs in the network; the larger the *ADP* is, the better the protocol performs. *OFT* is the time at which all the vehicles in the network finish downloading the file; the smaller the *OFT* is, the better the protocol performs.

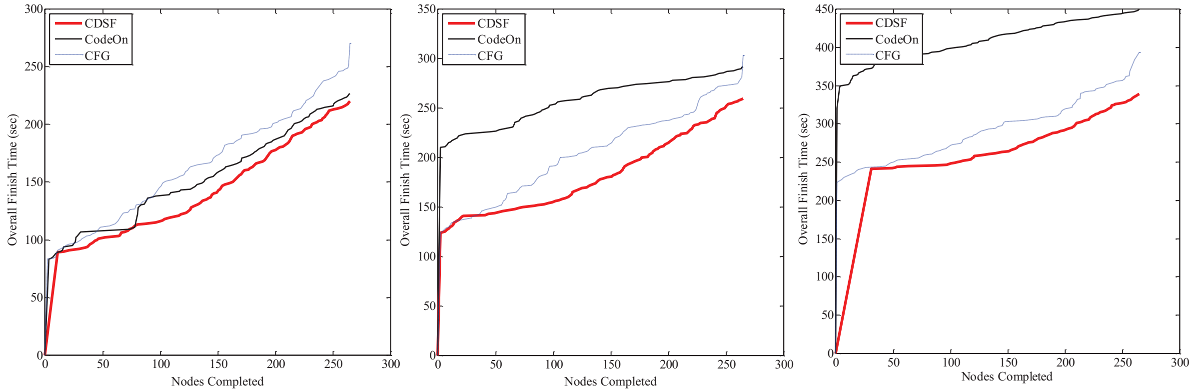
4.3.1. Downloading the popular content of different sizes

First we set the speed of the vehicles at 40 km/h and compare those protocols in downloading different size files.

Fig. 9 shows the average downloading percentage of the three protocols during the downloading process when the sizes of the files being downloaded are 64, 128 and 256 MB. The *x* axis is the simulation time, and the *y* axis is the *ADP*. Fig. 10 shows the overall finish time when different numbers of OBUs acquire the complete content. The *x* axis is the number of nodes completed, and the *y* axis is the *OFT*.

For the case of downloading a small-sized popular content, e.g. a 64 MB audio piece, ECDS performs best, as it takes the shortest time among all of the OBUs in acquiring the complete file. In the beginning, it may have a worse average downloading percentage than the comparison protocols. In this period, generations are accumulated to the core cells, and as the number of core cells is small, the average downloading percentage may be low. However, when the core cells accumulate sufficient generations, the large amount of non-core cells can quickly acquire the missing generations from the core cells in a few hops, thus the scope of the curve increases and the average downloading percentage surpasses those of the comparison methods. Network coding-based solutions (e.g. CodeOn) are as competitive as ECDS in this case, as they take less time among all of the OBUs to download 80% of the content.

When downloading a larger popular content, e.g. a 128MB advertisement clip, the advantage of ECDS becomes



a) Content size = 64MB
b) Content size = 128MB
c) Content size = 256MB

Fig. 10. The Overall Finish Time (OFT) in downloading the contents of different sizes.

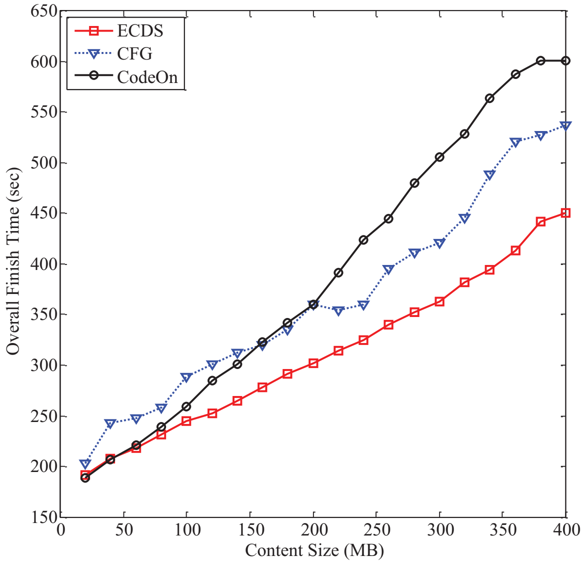


Fig. 11. The tendency of OFT in downloading different sized content.

distinct. It can be seen from Fig. 9(b) that the curve representing the downloading process of ECDS has the highest slope, which means that ECDS has the fastest downloading speed. Fig. 10(b) shows that ECDS takes the shortest time to complete the PCD process. As the requested number of decoders increases, the network coding based solution, i.e. CodeOn, becomes less efficient than downloading small files.

When downloading a much larger file, e.g. a 256MB compressed film, the advantage of ECDS is distinct, as it takes almost 1 min less to download the complete file than the second efficient protocol CFG.

Moreover, we find that ECDS performs superiorly to the comparison schemes when downloading larger content. Fig. 11 shows the comparison results of the time when all of the vehicles finish downloading different sized contents using the three PCD schemes. The results clearly show the following conclusions.

According to the above analysis, we draw the conclusion that ECDS performs well in downloading different sized contents in an urban vehicular network, especially in downloading large sized content.

4.3.2. Downloading content with different vehicle speeds

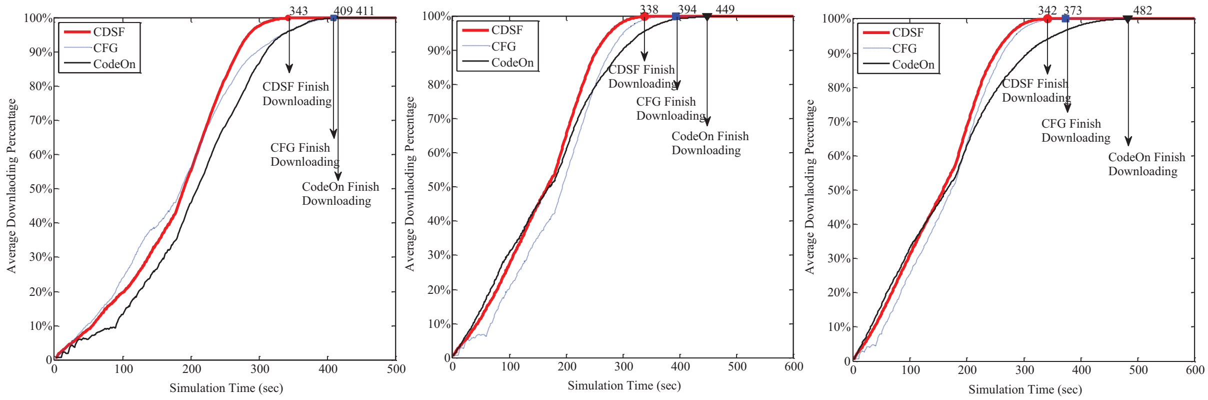
Second, we set the file size to 256MB and compared the protocols when the vehicles are moving at different speeds, e.g. 24, 40 and 60 km/h.

Fig. 12 shows the average downloading percentages of the three protocols during the downloading process when the vehicles are moving at different speeds, e.g. 24, 40 and 60 km/h. The x axis is the simulation time, and the y axis is the *ADP*. Fig. 13 shows the overall finish times when different numbers of OBUs acquire the complete content. The x axis is the number of nodes completed, and the y axis is the *OFT*.

It can be seen that the speed has little influence on the performance of ECDS. In all cases, ECDS has the highest average downloading percentage and the lowest overall finish time. It should be mentioned that CFG obtains better performance as the speed increases, as CFG is designed for freeway collaborative downloading. When the speed of the vehicles is excessively fast, the fundamental assumption in the topology pre-creation and periodical update scheme, i.e. the cell densities remain unchanged during a short interval, no more set up. As a result, it may be less efficient than CFG. As in a majority of cities in China, traffic speeds are restricted to 60 km/h, thus ECDS is of the best performance for PCD in urban vehicular networks.

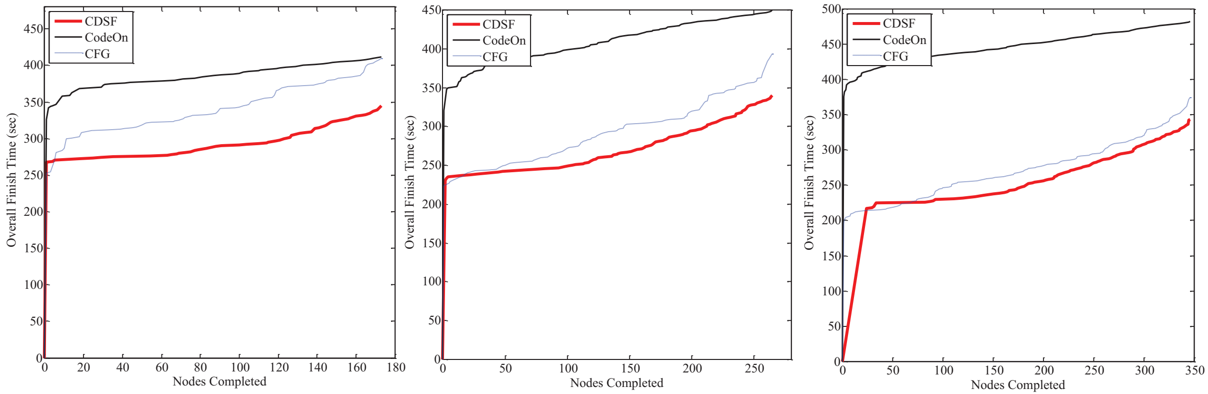
5. Conclusion

In this paper, we address the PCD problems in urban vehicular networks. Due to the high speed and limited bandwidth, the OBUs may fail in downloading the entire popular content, except for a few generations when passing through an RSU. When the OBUs are outside the coverage of the RSUs, they form a P2P network and collaboratively exchange the generations to complete the popular content dissimilation. For this, we propose



a) Speed = 24km/h b) Speed = 40km/h c) Speed = 60km/h

Fig. 12. The Average Downloading Percentage (*ADP*) in downloading 256 M content with different vehicles' speeds.



a) Speed = 24km/h b) Speed = 40km/h c) Speed = 60km/h

Fig. 13. The Overall Finish Time (OFT) in downloading 256 M content with different vehicles' speeds.

ECDS: an efficient collaborative downloading scheme for popular content distribution in urban vehicular networks. A cell-based clustering scheme is proposed, in which the OBUs join in different cells according to their geographical positions. By treating the cells as nodes, the positions of the nodes are fixed, which greatly simplifies the modeling of the VANETs. With the cell-based clustering scheme, a topology pre-creation and periodical update scheme is proposed to build the VANET of scale-free topology, which is beneficial to data dissemination. In the created topology, new relay selection and generation selection strategies are proposed, and the generations are first accumulated to several core cells, then quickly distributed to a majority of low-degree cells within a few hops, which accelerates the popular content distribution. The simulation results in multiple scenarios compared to two up-to-date collaborative PCD protocols, demonstrating the high performance of ECDS.

Acknowledgments

This research is supported by the National Natural Science Foundation of China Grants (61472001, U1405255) and the Local Governmental Foundation Grant (GY2013030, Zhenjiang).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.comnet.2016.02.006.

References

[1] S. Panwai, H. Dia, Comparative evaluation of microscopic car-following behavior, IEEE Trans. Intell. Transp. Syst. 6 (3) (2005) 314–325.

[2] M.H. Firooz, S. Roy, Collaborative downloading in VANET using network coding, in: Proceedings of the IEEE International Conference on Communications (ICC'12), Ottawa, ON, 2012, pp. 4584–4588.

[3] K. Mershad, H. Artail, A framework for secure and efficient data acquisition in vehicular ad hoc networks, IEEE Trans. Veh. Technol. 62 (2) (2013) 536–551.

[4] H. Zhu, R. Lu, X. Shen, X. Lin, Security in service-oriented vehicular networks, Wirel. Commun. 16 (4) (2009) 16–22.

[5] M. Gerla, L. Kleinrock, Vehicular networks and the future of the mobile internet, Comput. Netw. 55 (2) (2011) 457–469.

[6] M. Li, Z. Yang, W. Lou, Codeon: Cooperative popular content distribution for vehicular networks using symbol level network coding, IEEE J. Sel. Areas Commun. 29 (1) (2011) 223–235.

[7] T. Wang, L. Song, Z. Han, B. Jiao, Dynamic popular content distribution in vehicular networks using coalition formation games, IEEE J. Sel. Areas Commun. 31 (9) (2013) 538–547.

[8] M. Gerla, E.K. Lee, G. Pau, U. Lee, Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds, in: Proceedings of the 2014 IEEE World Forum on Internet of Things (WF-IoT), Seoul, South Korea, 2014, pp. 241–246.

[9] IEEE Standard for Information Technology Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks-Specific Requirements; Part 11, Jul. 2010.

[10] Draft Standard for Wireless Access in Vehicular Environments-Security Services for Applications and Management Messages, IEEE P1609.2/D9.0, May 2011.

[11] W. Zhu, D.J. Li, W. Saad, Multiple vehicles collaborative data download protocol via network coding, IEEE Trans. Veh. Technol. 64 (4) (2015) 1607–1619.

[12] X. Shen, X. Cheng, L. Yang, R. Zhang, B. Jiao, Data dissemination in VANETs: A scheduling approach, IEEE Trans. Intell. Transp. Syst. 15 (5) (2014) 2213–2223.

[13] D. Qiu, R. Srikant, Modeling and performance analysis of BitTorrent-like peer-to-peer networks, in: Proceedings of the 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM), 34, 2004, pp. 367–378.

[14] O. Heckmann, A. Bock, A. Mauthe, R. Steinmetz, The eDonkey file-sharing network, GI Jahrestagung 2 (51) (2004) 224–228.

[15] A. Nandan, S. Das, G. Pau, M. Gerla, M.Y. Sanadidi, Co-operative downloading in vehicular ad-hoc wireless networks, in: Proceedings of the Second Annual Conference on Wireless on-demand Network System and Services (WONS' 05), St. Moritz, Switzerland, 2005, pp. 32–41.

[16] O. Trullols-Cruces, M. Fiore, J.M. Barcelo-Ordinas, Cooperative download in vehicular environments, IEEE Trans. Mob. Comput. 11 (4) (2012) 663–678.

[17] C.F. Daganzo, The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory, Transp. Res. Part B Methodol. 28 (4) (1994) 269–287.

[18] J. Liu, J. Wan, Q. Wang, P. Deng, K. Zhou, Y. Qiao, A survey on position-based routing for vehicular ad hoc networks, Telecommun. Syst. (2015) 1–16, doi:10.1007/s11235-015-9979-7.

[19] J. Liu, J. Wan, Q. Wang, D. Li, Y. Qiao, H. Cai, A novel energy-saving one-sided synchronous two-way ranging algorithm for vehicular positioning, Mob. Netw. Appl. 20 (5) (2015) 661–672.

[20] J. Wan, D. Zhang, S. Zhao, L. Yang, Context-aware vehicular cyber-physical systems with cloud support: Architecture, challenges, and solutions, Commun. Mag. IEEE 52 (8) (2014) 106–113.

[21] C. Campolo, A. Molinaro, Multichannel communications in vehicular ad hoc networks: a survey, Commun. Mag. IEEE 51 (5) (2013) 158–169.

[22] R.K. Merton, The Matthew effect in science, Science 159 (3810) (1968) 56–63.

[23] Albert-László Barabási, Scale-free networks: A decade and beyond, Science 325 (5939) (2009) 412–413.

[24] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86 (14) (2001) 3200.

[25] S. Ahmed, S. Kanhere, VANETCODE: network coding to enhance co-operative downloading in vehicular ad-hoc networks, in: Proceedings of the International Conference on Wireless Communications and Mobile Computing (IWCMC' 06), Vancouver, British Columbia, Canada, 2006, pp. 527–532.

[26] C. Wu, S. Ohzahata, T. Kato, Network coding assisted cooperative relay scheme for sender-oriented broadcast in VANETs, in: Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC' 13), Shanghai, China, 2013, pp. 1369–1374.

[27] M. Liu, G. Chen, The sharing at roadside: Vehicular content distribution using parked vehicles, in: Proceedings of the IEEE INFOCOM'12, Orlando, FL, 2012, pp. 2641–2645.

[28] M. Gerla, C. Wu, G. Pau, X. Zhu, Content distribution in VANETs, Veh. Commun. 1 (1) (2014) 3–12.

[29] J. Schleich, G. Danoy, B. Dorronsoro, P. Bouvry, An overlay approach for optimising small-world properties in VANETs, in: Proceedings of the Sixteenth European Conference, EvoApplications'13, Grupo, Spain, 2013, pp. 32–41.

[30] A. Banerjee, R. Agarwal, V. Gauthier, C.K. Yeo, H. Afifi, F.B.S. Lee, A self-organization framework for wireless ad hoc networks as small worlds, IEEE Trans. Veh. Technol. 61 (6) (2012) 2659–2673.

[31] Christine Shea, Behnam Hassanabadi, Shahrokh Valaee, Mobility-based clustering in VANETs using affinity propagation, in: Proceedings of the Twenty-eight IEEE Conference on Global Telecommunications (GLOBECOM'09), Honolulu, HI, 2009, pp. 1–6.

[32] A. Lipowski, D. Lipowska, Roulette-wheel selection via stochastic acceptance, Phys. A Stat. Mech. Appl. 391 (6) (2012) 2193–2196.

[33] A. Sumalee, R.X. Zhong, T.L. Pan, W.Y. Szeto, Stochastic cell transmission model (SCTM): A stochastic dynamic traffic model for traffic state surveillance and assignment, Transp. Res. Part B Methodol. 45 (3) (2011) 507–533.

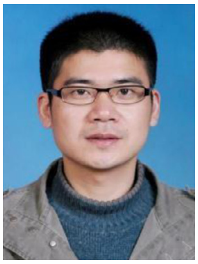
[34] J. Zhang, W. Mehner, S. Holl, M. Boltes, E. Andresen, A. Schadschneider, A. Seyfried, Universal flow-density relation of single-file bicycle, pedestrian and car motion, Phys. Lett. A 378 (44) (2014) 3274–3277.

[35] E. Moylan, D. Rey, S.T. Waller, Geometric congestion detection algorithms in the speed-flow and flow-density spaces, in: Proceedings of the Eighteenth IEEE International Conference on Intelligent Transportation Systems (ITSC), Chicago, USA, 2015, pp. 2763–2769.

[36] Pastor-Satorras, R., Castellano, C., Van Mieghem, P., and Vespignani, A. (2014). Epidemic processes in complex networks. doi:10.1103/RevModPhys.87.925 arXiv:1408.2701 [physics.soc-ph].

[37] F. Morone, H.A. Makse, Influence maximization in complex networks through optimal percolation, Nature 54 (2015) 65–68.

[38] J. Shao, S.V. Buldyrev, L.A. Braunstein, S. Havlin, H.E. Stanley, Structure of shells in complex networks, Phys. Rev. E 80 (3) (2009) 036105.



Liang-min Wang received his B.S. degree in Computational Mathematics in Jilin University, Changchun, China, in 1999, and the Ph.D. degree in Cryptology from Xidian University, Xi'an, China, in 2007. He is a full professor in the School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China. He has been honored as a “Wan-Jiang Scholar” of Anhui Province since Nov. 2013. Now his research interests include security protocols and Internet of Things. He has published over 60 technical papers at premium international journals and conferences, like IEEE Transactions on Vehicular Technology, IEEE GlobeCOM, IEEE WCNC. He has been severed as the TPC of many IEEE conferences, such as IEEE ICC, IEEE HPCC, IEEE TrustCOM. Now he is an associate editor of Security and Communication Networks, a member of IEEE, and a senior member of Chinese Computer Federation.

Wei Huang received his B.S. degree in Computer Science in Jiangsu University in the year of 2008. Now he is a postgraduate student of IoT department, Jiangsu University. He is also a member of the Portal Search Research, Baidu (China) Co., Ltd. His research interests include optimizing algorithm, graph theory and complex network theory. He has published one paper in IEEE iThings.