Cooperative transmission scheme based on mobility clustering for popular content distribution

Chen Guo*, Mohan Su
China Mobile Group Device Co., Ltd., Beijing 100053, China

ABSTRACT

Popular Content Distribution (PCD) is one of the main technologies to mobile communication. As the number of mobile nodes and popular content continue to increase, it is difficult to receive complete popular content when passing through a base station. In this paper, a cooperative transmission scheme based on mobility clustering is proposed, which aims to improve the completion rate of content and overall finish time. The simulation results show that for different sizes of contents, compared with ECDS and non-cooperative schemes, the proposed scheme reduces the overall finish time by 38% and 50% respectively.

Keywords: Popular content distribution, mobility clustering, cooperative transmission

1. INTRODUCTION

With the rapid development of the automobile industry, the number of mobile devices continues to increase. According to the survey of the International Organization of Motor Device Manufacturers (OMCA), by the end of 2023, the number of global mobile devices will exceed 1 billion, with the annual growth rate was 6.4%[1]. 95% of the mobile devices can be connected to the Internet by 2030. Some applications publish popular contents in the Area of Interest (AoI) through base station, such as popular videos, advertisements of nearby businesses, tourism information, real-time digital maps and real-time media content[2,3]. This service is called Popular Content Distribution (PCD) which has become a hot topic recently. As a result, the popular contents usually need to be provided to all mobile devices to maximize the delivery ratio. It is reported that the popular contents generally are large files. However, because the coverage of base stations is usually limited and the quality of wireless link is affected by devices mobility[4], it is difficult to receive complete popular content in the coverage of base stations. Once devices drive beyond the coverage of base stations, the content distribution will be interrupted. Because of the universal coverage, cellular networks attract lots of users and can carry higher demands for multimedia content services. However, the cellular networks-based solutions have disadvantages in the high communication cost and face the challenges of limited bandwidth and limited capacity caused by large number of devices[5]. In addition, the schemes based on DSRC suffer from the high cost of base station deployment.

Aiming to solve the problems of low delivery ratio and high delay of popular content distribution, we propose a cooperative transmission scheme based on mobility clustering for popular content distribution. Base stations encode the data packets by fountain code, and then send the encoded packets to the mobile devices in the coverage of base stations. according to the information of the mobile devices, base stations or base stations predict the mobile devices speeds by the BP neural network optimized by genetic algorithm. Based on the mobile device speeds, link quality and the differences of encoded packets, the mobile devices are divided into different clusters by using the unsupervised clustering algorithm. In each cluster, one or more mobile devices are selected as the senders according to the competition values of the mobile devices. The senders broadcast different encoded packets in the cluster in turn, and other mobile devices receive the broadcast packets as the receivers. After the cooperative transmission in each cluster, each mobile device can get enough encoded packets to decode. Because of the characteristics of fountain code, the Mobile devices can reconstruct the popular contents without knowing which packets are missing.

2. RELATED WORK

The idea of collaborate popular content distribution was first proposed by Alok Nandan et al. in the famous SPAWN protocol[6]. In the protocol, the Mobile devices always first request the rarest content from the nearest neighbor, which is

*guochen@cmdc.chinamobile.com

International Conference on Optics, Electronics, and Communication Engineering (OECE 2024), edited by Yang Yue, Proc. of SPIE Vol. 13395, 133954A · © 2024 SPIE · 0277-786X · Published

· doi: 10.1117/12.3049169

called “Rarest-Closet First” relay-selection strategy. In Reference[7], Trullols-cruces et al. considered the problem of using several road infrastructures and mobile devices to assist target mobile devices to download large files from road infrastructures in a more realistic complex road scenario. More recently, in References[8,9], the authors use coalition game to solve the problem of transmission in popular content distribution, which improves the average content percentage of content distribution, but it takes a lot of time in the coalition game stage. In Reference[10], trajectory prediction and coalition game theory are used to optimize popular content distribution, which effectively reduces the delay of PCD service. Zhang et al. proposed a distributed dynamic information transmission protocol for multiple target regions[11]. Wang et al. proposed a data distribution scheme based on directional clustering and probabilistic broadcast named as CPB[12]. In this scheme, different clusters are first construct among mobile devices according to the driving direction and geographical location of mobile devices. Wu et al. also proposed a cluster-based content distribution protocol which combines LTE (Long Term Evolution) with IEEE 802.11p for content distribution in VANETs[13]. In the process of clustering, a large amount of data is often forwarded among device nodes, which results in slow clustering speed, inaccurate clustering and high overhead. Higher clustering stability requires better clustering algorithms[14]. Intelligent algorithm is used in some papers for clustering. In Reference[15], Hafeez et al. proposed a distributed multi-channel mobile and aware cluster MAC (DMMAC) protocol. Through channel scheduling and an adaptive learning mechanism integrated in fuzzy Logic Inference System (FIS), mobile devices organize themselves into more stable and non-overlapped clusters. In Reference[16], Hassanabadi et al. proposed a clustering scheme based on mobility, which uses affinity propagation (AP) algorithm to cluster in a distributed manner. Bi et al. also proposed a global affinity propagation clustering (GAPC) algorithm based on mobility-related parameters and communication-related parameters[17]. Khan et al. proposed a novel cluster-based evolving graph model, which is suitable for the highly dynamic topology of mobile devices by using spectral clustering[18].

3. CLUSTER BASED ON MOBILITY

Considering the large number of mobile devices, cooperative transmission is complex and time-consuming, so it is impossible for all mobile devices to cooperate to share encoded packets in a same block. Therefore, the mobile devices that are beyond the coverage of base stations are split to several clusters, and then the mobile devices in the same cluster share data packets through cooperative transmission. Due to the dynamicity of mobile devices, the topology of mobile devices changes frequently. Spectral clustering is an excellent classifier to deal with the instantaneous shape of mobile devices. Mobile devices are divided into optimal number of groups by spectral clustering. In this section, we will focus on how to cluster mobile devices to improve the efficiency of cooperative transmission.

3.1 Affinity/similarity matrix

The affinity matrix describes the possibility of clustering two mobile devices. In this paper, we let link reliability and difference of encoded packets denote the similarity between two mobile devices.

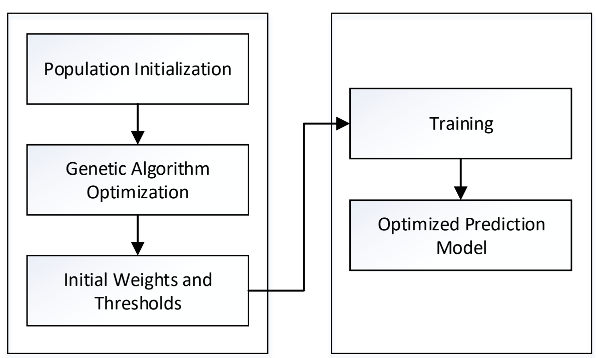


Figure 1. BP neural network optimized by genetic algorithm.

We can predict the future traveling speed of the mobile devices and the link existent time between the two mobile devices. T_{ij} denotes link existent time between mobile devices v_i and v_j. The device speed in the next 60s through the optimized BP neural network (As shown in Figure 1). To avoid too many output nodes, we take every 5 seconds as a period, and the average speed in this period represents the speed in this period. $v(t)$ denotes the speed of device v_i as below.

$$v(t) = \begin{cases} s_1, & 0 < t \le 5 \\ & \cdots \\ s_n, & 5 \times (n-1) < t \le 5 \times n \\ & \cdots \\ s_{12}, & 55 < t \le 60 \end{cases} \quad (1)$$

$\Delta v(\mathrm{t})_{ij}$ is the relative velocity between v_i and v_j. $\Delta V(\mathrm{t})_{ij}$ is the original function of $\Delta v(\mathrm{t})_{ij}$. T_{ij} can be calculated by

$$\int_0^{T_{ij}} \Delta v(t)_{ij}\, dt = R_{ij} \quad (2)$$

$$\Delta V(T_{ij})_{ij} = R_{ij} \quad (3)$$

where R_{ij} is the relative transmission range between device v_i and v_j. R_{ij} is related to the direction and speed of the device. When device v_i and v_j move in the same direction and v_i is in front, R_{ij} is:

$$R_{ij} = \begin{cases} R + d, & \textit{the average speed of } v_i \\ & \textit{is greater than that of } v_j \\ R - d, & \textit{the average speed of } v_j \\ & \textit{is greater than that of } v_i \end{cases} \quad (4)$$

when v_i and v_j move in opposite direction, R_{ij} is:

$$R_{ij} = \begin{cases} R - d, & \textit{if } v_i \textit{ and } v_j \textit{ moving away} \\ R + d, & \textit{if } v_i \textit{ and } v_j \textit{ moving to each other} \end{cases} \quad (5)$$

where d is the Euclidean distance between v_i and v_j as follows.

$$d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (6)$$

T_{ij} does not fully reflect the link reliability between v_i and v_j, so receive power is introduced into our scheme. Due to v_i and v_j are equipped with the same receiver and transmitter, and the channel environment is the same, so it is assumed that the receive power of v_i and v_j is the same, i.e., $P_{i,j}$=$P_{j,i}$. Link existent time T_{ij} and receive power $P_{i,j}$ jointly denote the link reliability between device v_i and v_j as below.

$$re_{ji} = re_{ij} = T_{ij} \times P_{i,j} \quad (7)$$

Mobile devices with high link reliability should be clustered into the same cluster to ensure the quality of communication. In addition to link reliability, the difference of encoded packets occupied by mobile devices is also an important factor to be considered when clustering mobile devices. Because Base stations send encoded packets to mobile devices in its coverage area by broadcast way, the similarity of encoded packets occupied by the mobile devices close to each other is very high. If only considering the geographical location in the traditional way and then the mobile devices close to each other in the geographical space are divided into the same cluster, there are usually few different encoded packets that can be shared with other mobile devices, so it is difficult to obtain enough encoded packets quickly through cooperative transmission for the mobile devices. Therefore, the difference of encoded packets occupied by the mobile devices should be also considered about the link reliability. $diff_{ij}$ denotes the difference of encoded packets occupied by v_i and v_j which is also equal to $diff_{ji}$ as below.

$$diff_{ij} = diff_{ji} = \frac{(P \backslash P_i) \cap P_j}{P} + \frac{(P \backslash P_j) \cap P_i}{P} = \frac{(P_j \backslash P_i) + (P_j \backslash P_i)}{P} \quad (8)$$

where P_i is the encoded packets owned by device v_i, P_j is the encoded packets owned by device v_j, and P is all encoded packets. Mobile devices with large differences of encoded packets should be clustered together. We use unsupervised spectral clustering algorithm to cluster the mobile devices on the road. Before clustering, it is necessary to determine the similarity between device nodes, which is the basis for clustering device.a_{ij} denotes the similarity between device v_i and v_j as below.

$$a_{ij} = re_{ij} \times diff_{ij} \quad (9)$$

where re_{ij} is the link reliability and $diff_{ij}$ is the difference of encoded packets occupied by device v_i and v_j. Matrix A denotes the similarity between all mobile devices in the network as below.

$$A = \begin{cases} a_{ij}, & if\ i \neq j \\ 0, & if\ i = j \end{cases} \tag{10}$$

A is called Affinity matrix having link reliability and encoded packet difference of n mobile devices based on Equation (11) as given below.

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \tag{11}$$

3.2 Cluster methodology

In communication field, lots of machine learning algorithms are adopted for clustering in paper, but most of them have poor effects on clustering shape (non-convex) and high dimension of data. Spectral clustering is suitable for the rapidly changing topology, so it is used to cluster mobile devices on the road. After getting the Affinity matrix A, the main steps of spectral clustering are as follows.

(1) Calculation of the degree matrix D. For each device v_i, the degree d_i can be calculated which is defined as the sum of a_{ij} of all mobile devices connected with device v_i as below.

$$d_i = \sum_{j=1}^{n} a_{ij} \tag{12}$$

It can also be considered that the degree value d_i is the sum of all elements in row i or column i of matrix A. According to the definition of the degree value, we can get a diagonal matrix D of dimension $n \times n$ as below.

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \tag{13}$$

(2) Calculation of the Laplacian matrix L. Laplacian matrix is the main tool of spectral clustering and Laplacian matrix L is computed as below.

$$L = D - A \tag{14}$$

The normalized Laplacian matrix L_{nor} is expressed as (15)

$$L_{nor} = D^{-0.5} L D^{-0.5} \tag{15}$$

(3) Eigenvalue decomposition. The minimum k eigenvalues of the normalized Laplace matrix L_{nor} and their corresponding eigenvectors are calculated. Then the eigenvectors form the eigenmatrix F of dimension $n \times k$. Each row in eigenmatrix F is used as a sample to be clustered with the clustering algorithm (such as K-means and other clustering algorithms), and finally k clusters are obtained. In order to find an optimal number of clusters k[19], the eigenvalues λ_i $i = \{1,2,\cdots,m\}$ of the normalized Laplace matrix L_{nor} are sorted in ascending order[18] based on the eigengap heuristic method and k is as follows.

$$k = argmax(\lambda_{i+1} - \lambda_i) \tag{16}$$

Algorithm 1 Algorithm for selecting senders

$CV_x = \{cv_{x1}, \cdots, cv_{xy}\}$ is the set of competition values of all mobile devices in the cluster C_x
$C_x = \{v_{x1}, \cdots, v_{xy}\}$ is the set of all mobile devices in the cluster C_x
$PACKET_{send}$ is the union of encoded packets occupied by all senders
num_{ps} is the size of $PACKET_{send}$
num_{last} is the size of $PACKET_{send}$ in the last loop
num_{target} is the number of encoded packets required for decoding
Sort the mobile devices of C_x in descending order according to the values of CV_x
for v_{xi} **in** C_x
  add v_{xi} to S_x
  $PACKET_{send}$=$PACKET_{send} \cap P_{xi}$
  if $num_{ps} \geq num_{target}$ **or** $num_{ps} = num_{last}$
    break
  end if
  $num_{target} = num_{ps}$
end for

3.3 Selecting relay mobile devices

The mobile devices are in the same cluster, and the dashed circle represents the coverage area of the sender. In a cluster, when a sender broadcast encoded packets, there may be a few mobile devices outside the coverage range of the sender. So, it is necessary to select appropriate relay mobile devices to forward the broadcast packets to these mobile devices. Routing reliability is the main factor for selecting a relay device. For example, when v_{xa} wants to broadcast encoded packets in cluster C_x, if v_{xb} is beyond the coverage of v_{xa}, v_{xa} need to find mobile devices with high route reliability among its neighbors as relay nodes. $Route_{ab} = (v_{xa}, v_{xb})$ denotes the Route from v_{xa} to v_{xa} with k links: $l_1 = (v_{xa}, relay_1)$, $l_2 = (relay_1, relay_2), \cdots$, $l_k = (relay_1, relay_2)$. Each link l_1 has specific link reliability re_{l_i} that is calculated using Equation (17). Therefore, the reliability of $Route_{ab}$ is denoted by $r(Route_{ab})$ as below.

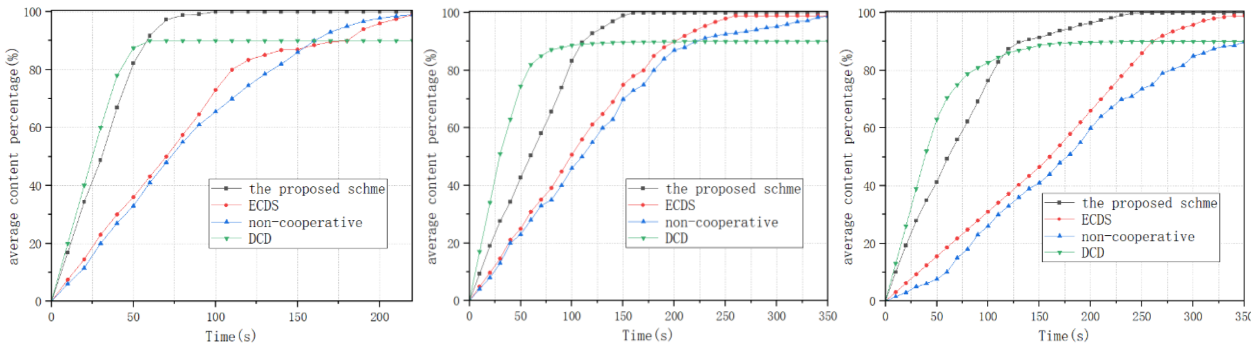
$$r(Route_{ab}) = \min(l_1, l_2, \cdots, l_k) \tag{17}$$

The route reliability is determined by the link with the least link reliability in the route. Among all the routes from v_{xa} to v_{xb}, the route with the highest reliability is selected for forwarding encoded packets to ensure that all mobile devices in the cluster can receive the encoded packets broadcast by the sender. Algorithm 1 describes how to select senders.

4. SIMULATION

In this section, the trained BP neural network model for predicting device speeds is used to predict the device speeds of the test set. Then, the performance of the proposed CVMC scheme is simulated and analyzed. The CVMC scheme is compared with other schemes: the ECDS scheme, a cooperative PCD scheme, the DCD scheme, and the non-cooperative scheme, a basic non-cooperative scheme of PCD. The comparisons illustrate the advantages of our CVMC scheme over non-cooperative schemes, and the performance improvement compared with other cooperative schemes. The non-cooperative scheme is that the device receives the data packets broadcast by the RSU in the V2R phase, and then randomly chooses a data packet to send to its neighbor mobile devices. The ECDS scheme is a cooperative download scheme for PCD using cell-based clustering and inter-relay selection. The DCD scheme[8] is a PCD scheme for cooperative device networks based on big data-based device trajectory prediction strategy and a coalition game based on resource allocation mechanism. Our simulation experiments are carried out on the MATLAB R2021b platform.

In Figure 2, we compare the average content percentage of the three schemes over time. The number of mobile devices is set to 200, the average speed is 20 m/s, and the sizes of the files are 64 MB, 128 MB, and 256 MB respectively. The x axis is the time and y axis is the ACP. It shows that the ACP of the proposed scheme can reach 100% faster. For the sizes of the files are 64 MB, 128 MB and 256 MB respectively, when the average content percentage increases from 0 to 90%, the time taken by the proposed scheme only is about 1/2 of that of the ECDS scheme and only 3/8, 1/2 and 1/3 of that of the non-cooperative scheme respectively. In the non-cooperative scheme, channel contention and repeated transmission occur when multiple mobile devices broadcast. The ECDS scheme improves ADP over non-cooperative schemes by allowing cooperative transmission. The proposed scheme in this paper adopts fountain code. Due to the random coding and large sizes of distributed files, it is almost impossible that different RSUs can broadcast the same encoded packets. Therefore, the problem of repeated transmission is avoided and the broadcast efficiency is improved. The vehicles also don't have to sort the received encoded packets. Although DCD performs best at first, it quickly reaches the ceiling about 90% ACP. In DCD, only the OBUs that complete the popular content distribution can transmit data as the senders, which leads to the performance ceiling of DCD



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

Figure 2. Comparison between different schemes in terms of average content percentage with different content sizes.

5. CONCLUSION

In this paper, we have proposed a cooperative transmission scheme based on mobility clustering to solve the PCD problem. Due to large file size, fast device speed and limited bandwidth, the mobile devices cannot receive complete content in the coverage of base stations. Thus, the mobile devices should cooperate with neighbors when they beyond the coverage of Base stations. In the proposed scheme, base stations encode the large files with fountain codes. Mobile devices only need to receive a certain number of encoded packets to decode because of the characteristics of fountain code. The simulation results show that compared with the non-cooperative scheme and ECDS scheme, the CVMC scheme accelerates the content distribution, and performs better in terms of ACP and OFT of mobile devices in multiple scenarios.

REFERENCES

[1] OICA, "International organization of motor device manufacturers," <https://www.oica.net/> (20 May 2021).
[2] Chen, C., Hu, J., Qiu, T., Atiquzzaman, M. and Ren, Z., "CVCG: Cooperative transmission scheme based on coalitional game for popular content distribution in vehicular ad-hoc networks," IEEE Transactions on Mobile Computing 18(12), 2811-2828 (2019).
[3] Lu, J., Yang, W. and Wu, F., "High-Definition map distribution in named data networking based VANETs," 2020 3rd International Conference on Hot Information-Centric Networking (HotICN), 129-134 (2020).
[4] Guo, X., Chen, Y., Cao, L., Zhang, D. and Jiang, Y., "A receiver-forwarding decision scheme based on Bayesian for NDN-VANET," China Communications 17, 106-120 (2020).
[5] Huang, W. and Wang, L., "ECDS: Efficient collaborative downloading scheme for popular content distribution in urban vehicular networks," Computer Networks 101, 90-103 (2016).
[6] Nandan, A., Das, S., Pau, G., et al., "Cooperative downloading in vehicular ad-hoc wireless networks," Second Annual Conference on Wireless On-Demand Network Systems and Services, 32-41 (2005).
[7] Trullos, O., Morillo-Pozo, J., Barcelo, J. M., et al., "A cooperative vehicular network framework," Proceedings of the 20th International Conference, 2807-2812 (2009).

[8] Hu, J., Chen, C. and Liu, L., "Popular content distribution scheme with cooperative transmission based on coalitional game in VANETs," 2018 21st International Symposium on Wireless Personal Multimedia Communications (WPMC), 69-74 (2018).
[9] Fei, X., Luan, X., Yi, N., et al., "Content distribution in vehicular networks using coalitional graph game," 16th International Conference on Advanced Communication Technology, 1194-1197 (2014).
[10] Zhou, Z., Yu, H., Xu, C., et al., "Dependable content distribution in D2D-based cooperative vehicular networks: A big data-integrated coalition game approach," IEEE Transactions on Intelligent Transportation Systems 19(3), 953-964 (2018).
[11] Zhang, L., Gao, D., Gao, S. and Leung, V. C. M., "Smartgeocast: dynamic abnormal traffic information dissemination to multiple regions in VANET," Wireless Communications and Mobile Computing Conference, 1750-1755 (2013).
[12] Liu, L., Chen, C., Qiu, T., Zhang, M., Li, S. and Zhou, B., "A data dissemination scheme based on clustering and probabilistic broadcasting in VANETs," Vehicular Communications 13, 78-88 (2018).
[13] Wu, C., Chen, X., Zhang, L. and Ji, Y., "Cluster-based content distribution integrating LTE and IEEE 802.11p with fuzzy logic and Q-learning," IEEE Computational Intelligence Magazine 13, 41-50 (2018).
[14] Zhang, J., Ren, M. Y. and Labiod, H., "Performance evaluation of link metrics in device networks: A study from the cologne case," Proc. 6th ACM Symp. Develop. Anal. Intell. Veh. Netw. Appl. (DIVANet), New York, NY, USA, 123-130 (2016).
[15] Hafeez, K. A., Zhao, L., Mark, J. W., Shen, X. and Niu, Z., "Distributed multichannel and mobility-aware cluster-based mac protocol for vehicular ad hoc networks," IEEE Trans. Veh. Technol. 62(8), 3886-3902 (2013).
[16] Hassanabadi, B., Shea, C., Zhang, L. and Valaee, S., "Clustering in vehicular ad hoc networks using affinity propagation," Ad. Hoc. Netw. 13(1), 535-548 (2014).
[17] Bi, X., Guo, B., Shi, L., Lu, Y., Feng, L. and Lyu, Z., "A new affinity propagation clustering algorithm for V2V-supported VANETs," IEEE Access 8, 71405-71421 (2020),
[18] Khan, Z., Fan, P., Fang, S. and Abbas, F., "An unsupervised cluster-based VANET-oriented evolving graph (CVoEG) model and associated reliable routing scheme," IEEE Transactions on Intelligent Transportation Systems 20, 3844-3859 (2019).
[19] Liu, J., Xue, K., Miao, Q., Li, S., Cui, X., Wang, D. and Li, K., "MCVCO: Multi-MEC cooperative vehicular computation offloading," IEEE Transactions on Intelligent Vehicles 9, 813-826 (2024)