IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 25, 2020, accepted June 3, 2020, date of publication June 9, 2020, date of current version July 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3001140

Dynamic Resource Scheduling Optimization With Network Coding for Multi-User Services in the Internet of Vehicles

CHEN HUANG[1,2], JIANNONG CAO[3], (Fellow, IEEE), SHIHUI WANG[1,2], AND YAN ZHANG[1,2]

[1]School of Computer Science and Information Engineering, Hubei University, Wuhan 430062, China
[2]Hubei Engineering Research Center of Educational Informationalization, Wuhan 430062, China
[3]Department of Computing, The Hong Kong Polytechnic University, Hong Kong

Corresponding author: Chen Huang (huang@hubu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61977021, in part by the Natural Science Foundation of Hubei province under Grant 2018CFB692, in part by the Science and Technology Innovation Program of Hubei Province under Grant 2018ACA13, and in part by the Science and Technology Innovation Major Program of Hubei Province under Grant 2019ACA144.

ABSTRACT For Internet of Vehicles (IoV) systems with multiple users, network coding can be introduced to provide efficient error control and throughput improvement services. However, if the heterogeneity characteristics and requirements of the end users (vehicles) are neglected, it will be difficult for an IoV system to provide each end user with fair system services, without which the advantages of network coding cannot be fully achieved and the performance of the multi-user diversity system will be degraded. In this paper, we propose a Dynamic Resource Scheduling Optimization (DRSO) algorithm, a dynamic fair scheduling algorithm combined with network coding for system resource allocation in a multi-user IoV system. We construct a general solution framework for service scheduling: first, we estimate the fairness index for each end user (vehicle) with the key information on Quality of Service (QoS). Second, we construct a service scheduling control model based on the service capability of control entities (multi-access edge computing servers), and propose a new utility evaluation function. Third, based on the fairness index, we select end users into multiple network coding sets. Network coding sets are the basic units of service scheduling. The optimization objective of the scheduling service is to maximize the total utility of all the network coding sets (the utility of the control entity). Finally, we establish a coding cache queue in the control entity based on the scheduling decision. To obtain the global optimal solution for active queue control, we combine a Quantum Particle Swarm Optimization (QPSO) algorithm with a Proportional Integral (PI) model. Then, the optimal scheduling decision can be made. Extensive simulation results show that DRSO outperforms related scheduling algorithms in varying traffic loads, demonstrating that DRSO can effectively guide service resource allocation.

INDEX TERMS Multi-user, fairness control, network coding set, cache queue, internet of vehicles.

I. INTRODUCTION

As the infrastructure of mobile vehicle networks, the Internet of Vehicles (IoV) combines the advantages of the Internet of Things (IoT) and an Intelligent Transportation System (ITS). The IoV is a multi-source, multi-destination, and multi-user wireless network system with the characteristics of unstable network topology, fast node mobility, and frequent data exchange. End users (vehicles) share the wireless channel resource (accessing service time) allocated by the control entity (multi-access edge computing server) to obtain more transmission opportunities.

The associate editor coordinating the review of this manuscript and approving it for publication was Yong Yang.

Due to the instability of the wireless channels, the IoV may encounter channel error during transmission, and the network throughput is constrained. For error control and throughput improvement, network coding can be introduced due to its advantage in data fusion [1]. Intermediate nodes conduct linear network coding on multiple original data packets to generate the coded data packets. The destination recovers the original data packets according to the coding matrix.

Network coding can effectively reduce retransmission, and the theoretical limit of network throughput in multicast can be achieved [2].

However, traditional random linear network coding does not consider the diversity of end users in the IoV. On the one hand, vehicles vary in link reliability and service requirements in a complicated traffic environment. Without being allocated accessing service time, some end users may never be able to take part in network coding. On the other hand, if the service requirements from an end user are satisfied without the consideration of its channel quality, the performance of other end users may be degraded, this is the well-known ''crying baby'' problem in a multi-user network system [3]. Literature [4] has also pointed out that, the advantage of network coding in throughput improvement can only be fully achieved by proper scheduling. A fair accessing service time, network throughput and cache cost should be considered together to make an optimal scheduling decision in a multi-user IoV system [5]. Otherwise, the IoV system may suffer from decreased throughput and increased transmission delay when the channel quality is poor [6].

Network coding focuses on improving the overall network throughput without considering fair service resource allocation. Multi-user scheduling emphasizes providing fair service to each end user. Therefore, the combination of multi-user scheduling technology and network coding can be of great help to improve the channel gain and achieve higher network throughput for the IoV system.

To address the shortcomings of prior efforts, we propose a dynamic resource scheduling optimization algorithm referred to as DRSO, which provides fair network access and error control services in a multi-user IoV system. This algorithm gives comprehensive consideration to the various factors affecting service scheduling and quantifies these influencing factors. We construct a general solution framework according to DRSO. **First**, the control entity perceives the channel and service state of the IoV system in a real-time manner according to the feedback messages from end users. Given the key attribute information on quality of service (QoS) requirements, we make the initial measurement of the fairness index of an end user's service requirement. **Second**, utilizing the characteristics of multi-user diversity, a new service scheduling control function for utility estimation is constructed based on the service capability of the control entity. We calculate the service utility for the end user by considering more complicated factors influencing the service scheduling decision. **Third**, based on the fairness index, end users are selected into network coding sets. The network coding sets are the basic unit of service scheduling and are managed by the control entity. The total utility of network coding sets (the utility of the control entity)is maximized as the optimization objective of service scheduling. Then the weights of the service's key attributes can be determined. **Finally**, we make the scheduling decision and establish a coding cache queue in the control entity. A Quantum Particle Swarm Optimization (QPSO) algorithm is combined with a Proportional Integral (PI) model to solve the global optimal solution of active queue control. The accessing service time is fairly allocated to each end user.

This paper makes the following contributions:

1) We formally formulate the multi-user accessing service problem and devise a multi-layer solution framework for service scheduling between end users and control entity in the IoV system. (Section III)
2) We construct a statistical fairness measurement model with the concept of the fairness index, which can ensure the independence of end users. In addition, we propose a new credit-based service scheduling control model to evaluate the service utility of end users. (Section IV)
3) We design an improved network coding method for constructing a network coding set as the basic unit of service scheduling. An optimization algorithm for maximizing the total utility of network coding sets is proposed for determining the weights of the service's key attributes. (Section V)
4) We propose a coding cache queue control algorithm. This algorithm solves the global optimal solution of active queue control, and the caching overhead of network coding can be constrained. (Section V)

The remainder of this paper is organized as follows: we review related work in Section II. Section VI demonstrates simulation results, and Section VII presents the conclusions and future work of this paper.

II. RELATED WORKS

The multi-user diversity technique based on a scheduling strategy has been broadly used for the high-speed transmission of data packets in wireless cellular systems [7]. Temporal fairness scheduling is the main focus in recent studies. Classic multi-user systems such as the High Data Rate (HDR) system and High-Speed Data Packet Access (HSDPA) system, can obtain the multi-user diversity gain through Time-Division Multiple Access (TDMA) [8]. Shahsavari *et al.* [9] propose activating specific users to determine the set of feasible temporal shares in nonorthogonal multiple access system. In this paper, we also discuss multi-user diversity in the time domain.

There are two scheduling directions in a wireless network: uplink and downlink. For uplink scheduling, to alleviate packet loss and improve resource allocation, Ferng *et al.* [10] propose allowing the user equipment with the least left delay budget to be scheduled first. For downlink scheduling, Liu *et al.* [11] propose a channel-aware scheduling scheme to exploit both spatial and multi-hop diversity. The scheduling decisions are made at every time slot based on the designed priority indexes.

To approximate the optimal scheduling solution, a heuristic algorithm with polynomial time is investigated. In a centralized time-slotted channel hopping network, a scheduling algorithm for maximizing the network throughput is proposed [12]. Li *et al.* [13] also propose a heuristic algorithm for an energy harvesting mobile sensor network under the constraints of radio link quality while ensuring fair data reception. To maximize the aggregate throughput subject to a fairness constraint, Ge *et al.* [14] optimize the selection of user sets to approximate a nonconvex optimization problem. Proportional fair scheduling is a newly proposed method for energy efficiency. Liu *et al.* [15] present an analytical model for the performance analysis of dynamic proportional fair scheduling, and design a low-complexity algorithm for joint power allocation and user set selection. Li *et al.* [16] also propose normalized Signal-to-Noise Ratio (SNR) based fair scheduling for a multicast system.

To achieve throughput optimality with flow-level scheduling, Chen *et al.* [17] propose a scheduling algorithm for fairly scheduling TCP flows in wireless networks with time varying channel conditions. Velkov *et al.* [18] realize optimal proportional fair scheduling when each frame consists of either a downlink energy harvesting phase or an uplink information transmission phase. In a cellular system, Parruca *et al.* [19] present a closed-form analytical model for the throughput expectation of proportional fair scheduling, and the model takes into account a precise Signal-to-Interference-and-Noise Ratio (SINR) distribution. Furthermore, for the effective reuse of radio resources in a device-to-device communication system, Gu *et al.* [20] propose an optimal fair scheduling scheme that maximizes the logarithmic sum of the user data rates.

To improve the quality of services and enhance the resource allocation management in the IoV, Chen *et al.* [21] propose a virtual framework using a learning-based resource allocation scheme for mobile vehicle service. Huang *et al.* [22] also propose a service-oriented network architecture to reduce the traffic load and simplify network management with a service aggregation and caching (SAaC) scheme. Li *et al.* [23] propose a machine learning based code dissemination scheme to choose vehicles with higher reliable degree and coverage ratio to deliver code with lower costs.

III. DRSO DESCRIPTION AND PROBLEM DEFINITION

A. MULTI-ACCESS EDGE COMPUTING ARCHITECTURE

The DRSO algorithm proposed in this paper is used in a typical IoV system with multi-user access. The IoV system integrates various traffic information for frequent packet exchange among vehicle users. As shown in Fig. 1, we consider a Multi-access Edge Computing (MEC) architecture for various Vehicle to X (V2X) information services. MEC servers (control entities) are deployed on Road Side Units (RSUs), and provide system service resources for vehicles (end users). The control entities are connected with each other in an edge cloud, and the end users access the control entities through the wireless channels of the RSUs to obtain computational resources. Each control entity manages multiple end users within its transmission range. The control entities have moderate computing and cache capabilities, and can only serve one computational task at a time.

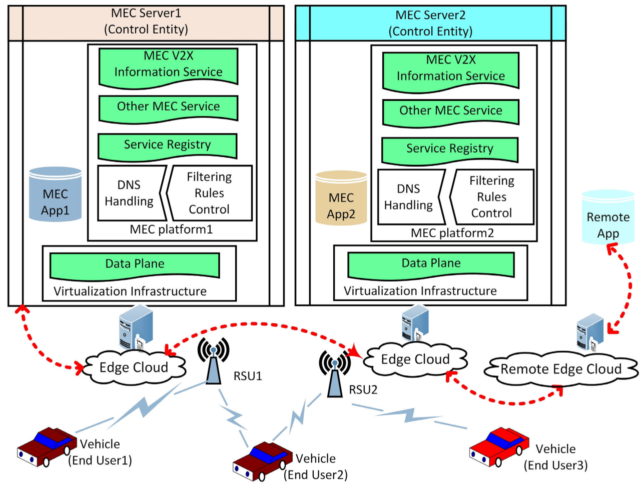


FIGURE 1. The architecture of IoV communication system.

B. WIRELESS CHANNEL MODEL AND MULTI-LAYER SOLUTION FRAMEWORK

In this paper, we adopt an explicit channel notification mechanism on the physical layer for querying Channel State Information (CSI) [24]. The uplink and downlink of the wireless channel are independent, and do not interfere with each other. In the interaction between the end users and a control entity, **first**, end users report CSI to the control entity dynamically through the uplink channel. Therefore, the detailed CSI of the physical layer can be used for the calculation of the transmission cost. **Second**, the IoV system estimates the channel state and evaluates the service capability of the control entities. **Third**, the service scheduling decision is made and distributed through the downlink channel to all the end users.

Table. 1 summarizes the frequently used notations and their meanings in this work.

Based on the above explicit channel notification mechanism, we obtain feedback data from end users to cover as many factors that affect the different kinds of IoV services as possible. Specifically, streaming media service is mainly affected by bandwidth and throughput; interactive service is mainly affected by the delay and packet loss rate; VoIP session service is mainly affected by the delay jitter and bit error rate; and non real-time best-effort service is mainly affected by the bandwidth and packet loss rate. The related factors for each service are detailed in Table. 2.

Fig.2 presents the multi-layer solution framework for system resource scheduling proposed in this paper. There are four layers in the framework. At the bottom Layer for the Service (**LS**), we receive all the feedback information from end users, and extract the key attributes of service requests; at the second Layer for the End User (**LEU**), we evaluate the fairness index in multiple forms according to the attributes of service requests; at the third Layer for the Network Coding Set (**LNCS**), we calculate the utility of a network coding set based on the utility of member end users; and at the top Layer for the Control Entity (**LCE**), we maximize the utility of the

TABLE 1. Definitions of relative parameters and acronyms.

Definition	Description
t	Time slice
$F_{I(k,i)}(t)$	Fairness index for end user k requesting service i
$R_{I(k)}(t)$	Instantaneous transmission rate of end user k
$\bar{R}_{(i)}(t-1)$	Average transmission rate of service i
$D_{I(k,i)}(t)$	Instant service delay for end user k requesting service i
$D_{M(k,i)}$	Maximum service delay for end user k requesting service i
C_F	Compensation factor
$GBR_{(i)}$	the value of GBR for the real-time service i
$F_{NI(k,i)}(t)$	Fairness index considering C_F based on $F_{I(k,i)}(t)$
$F_{SI(k,i)}(t)$	Statistical fairness index based on $F_{NI(k,i)}(t)$
$T_{A(k)}(t)$	The accessing service time allocated to end user k
$T_{R(k)}(t)$	The accessing service time that the end user k requests
G_X	Threshold for the members of a network coding set
$U_{E(k,i)}(t)$	The utility for end user k requesting for service i
$U_{(k)}(t)$	The utility for end user k
$U_S(t)$	The utility for the network coding set
U_C	The utility for the control entity
LS	Layer for the service
LEU	Layer for the end user
$LNCS$	Layer for the network coding set
LCE	Layer for the control entity
RSU	Road side unit
MEC	Multi-access edge computing

TABLE 2. Related factors of multiple services.

Service Type	Transmission Mode	Main Factors	Secondary Factors
Streaming media service	Unidirectional	Bandwidth Throughput	Delay Packet loss rate
Interactive service	Bidirectional	Bandwidth Throughput	Bandwidth Throughput
VoIP session service	Bidirectional	Delay jitter Delay	Bit error rate Throughput
Best effect service	Unidirectional	Bandwidth Bit error rate	Delay Throughput

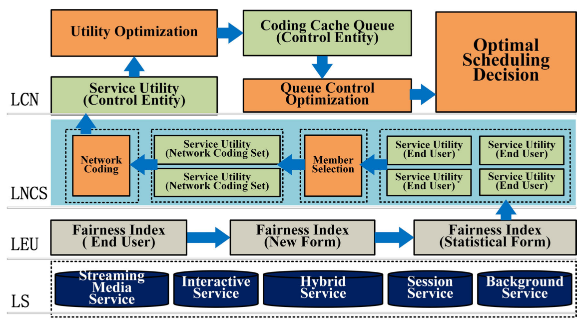


FIGURE 2. Solution framework for fair scheduling.

control entity, and optimize the coding cache queue to make the optimal scheduling decision.

IV. FAIRNESS MEASUREMENT AND SERVICE SCHEDULING CONTROL

In designing the solution framework for providing fair system service in an IoV system, it is important to measure the fairness for specific end user k with service i first. Generally, an accurate fairness measurement is an essential reference for estimating a single end user's service utility in service scheduling control. Therefore, in this section, **first**, in the layer for the service, we estimate the fairness index for an end user k requesting service i based on QoS related information. **Second**, in the layer for the end user, by utilizing the characteristics of multi-user diversity, we construct the service scheduling control model and calculate a single end user's service utility.

A. MEASUREMENT OF THE FAIRNESS INDEX (LAYER FOR THE SERVICE)

1) CALCULATION OF THE FAIRNESS INDEX

Once an end user makes a request for accessing service time, we initialize the procedure for fairness measurement. We introduce the concept of the fairness index from a Multi-User Multiple-Input-Multiple-Output (MU-MIMO) system [25]. The basic idea of the fairness index is balancing between fairness and throughput. In an IoV system, the fairness index is related to the multiple key attributes of QoS requirements from CSI, including the packet loss rate, average transmission rate, instantaneous transmission rate, instant service delay, and maximum service delay [26].

According to the feedback from the end users, the control entity can calculate the fairness index for each end user. If we do not consider the difference between real-time service and non-real-time service, at time slice t, the fairness index for end user k requesting service i can be defined as [26]:

$$F_{I(k,i)}(t) = -\lg(\delta_i) \cdot \frac{R_{I(k)}(t)}{\bar{R}_{(i)}(t-1)} \cdot \left[\frac{D_{I(k,i)}(t)}{D_{M(k,i)}}\right] \tag{1}$$

where, $\lg(\delta_i)$ represents the packet lose rate of service i. $R_{I(k)}(t)$ represents the instantaneous transmission rate of end user k. $\bar{R}_{(i)}(t-1)$ represents the average transmission rate of service i, and $D_{I(k,i)}(t)$ represents the instant service delay. $D_{M(k,i)}$ represents the maximum service delay.

However, in an IoV system, there are never surplus resources for non-real-time service because of the nonstop packet transmission. If the increased degree of non-real-time service is higher than that of real-time service, the fairness of real-time service will be compromised. Therefore, we introduce a compensation factor including the Guaranteed Bit Rate (GBR) [27] to the fairness index C_F to ensure the fairness of real-time service as follows:

$$C_F = \max\left(1, \frac{GBR_{(i)}}{\bar{R}_{(i)}(t-1)}\right) \tag{2}$$

where $GBR_{(i)}$ represents the value of the GBR for real-time service i. Based on Eq(1) and C_F, we construct the fairness index $F_{NI(k,i)}(t)$ considering compensation factor as follows:

$$F_{NI(k,i)}(t) = F_{I(k,i)}(t) \cdot C_F \tag{3}$$

If the average rate does not reach $GBR_{(i)}$, the value of the compensation factor C_F is greater than 1. The priority of the real-time service rises, and more system resources can be obtained in the following time slice.

2) STATISTICAL FAIRNESS INDEX

In a wired network, the Proportional Fairness Index (PFI) is used to describe the difference in throughput among the links [28]. However, the fairness of throughput cannot ensure the independence of end users. If a normalized throughput is adopted, an end user with poor channel quality may obtain an unfair channel accessing time, which will cause throughput decline in the whole system.

To adapt to the complicated environment of the IoV system, we make two modifications to the PFI: **First**, a statistical fairness index that can reflect the short-term fluctuation of wireless communication in the IoV is needed. End users with stable channel conditions can receive better system service. **Second**, we focus on the fairness of accessing service time, instead of the fairness of network throughput. The fairness of network access time can ensure the independence of end users, and the influence of end users with poor channel quality is controllable and predicable.

To distinguish the fairness of network access time from the fairness of network throughput, a Statistical Fairness Index (SFI) for accessing service time can be calculated based on Eq(3), and is defined as:

$$F_{SI(k,i)}(t) = \Pr\left(\left|\frac{F_{NI(k,i)}(t+\Delta t) - F_{NI(k,i)}(t)}{\Delta t}\right| \geq F_X\right) \tag{4}$$

where F_X represents the threshold of the SFI, and Δt represents the time interval.

B. SERVICE SCHEDULING CONTROL (LAYER FOR THE END USER)

In the layer for the end user, we construct a service scheduling control model for the allocation of accessing service time, as shown in Fig.3. Each end user has multiple types of service requests, such as streaming media service, interactive service, session service. The control entity manages the radio bearer resource for allocating time slices to the end users and caches the original data packets for the service requests.

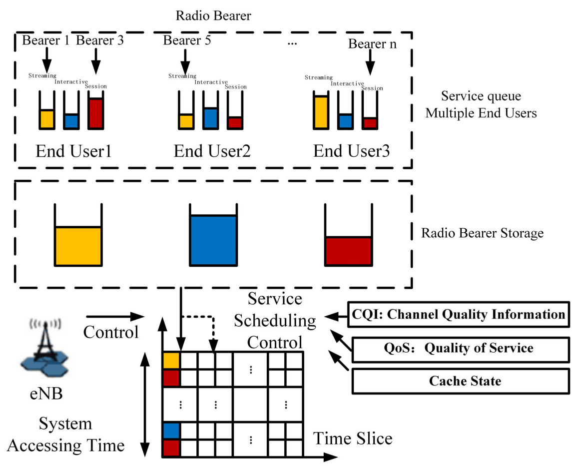


FIGURE 3. **Service Scheduling Control Model.**

First, we estimate the control entity's service capability based on a credit calculation method [29]. **Second**, we construct a service scheduling control model based on the control entity's service capability and the end user's fairness index, and propose a new utility estimation function.

1) EVALUATION OF SERVICE CAPABILITY

The system resources are limited, and the total accessing service time is constrained. Therefore, it is necessary to have a clear evaluation of the control entity's service capability. We introduce a credit calculation method used in bank transactions to describe the service capability [29]. A credit is a monetary unit deposited in a bank (control entity), the end user can obtain credit from the control entity according to the fairness index in Eq(4). The control entity's initial credits are equal to the total accessing service time that can be provided. For the control entity, more credits indicate a higher service capability; for the end user, more credits indicate that the end user should obtain more service resource. Therefore, credits can be used to balance service resources among the end users. The design of the credit updating algorithm is as follows:

Step 1: All the end users' credits are initialized to 0. The control entity has the same amount of initial credits as the total accessing service time that it can provide.

Step 2: The control entity allocates the credits among the end users based on each end user's fairness index. Then each end user updates its credits.

Step 3: The control entity allocates accessing service time to the end users according to each end user's credits.

Step 4: When the accessing service time for an end user is ended, the control entity restores and accumulates the accessing service time to provide further services. Go to Step 2.

At time slice t, we define $T_{R(k)}(t)$ as the accessing service time that end user k requests. $T_{A(k)}(t)$ represents the accessing service time allocated to end user k. $T_C(t)$ represents the total accessing service time that the control entity can provide. $C_{E(k)}(t)$ represents end user k's credits. $C_C(t)$ represents the control entity's credits. $C_{A(k)}(t)$ represents the credits that the control entity allocates to end user k.

To reveal the service capability of the control entity, we update the statistical fairness index in Eq(4), and propose a new form of fairness index $F_{CI(k,i)}(t)$ for end user k requesting service i considering the allocation of accessing time service, and $F_{CI(k,i)}(t)$ is shown as follows:

$$F_{CI(k,i)}(t) = F_{SI(k,i)}(t) \cdot \left(\frac{T_{A(k)}(t-1)}{T_{R(k)}(t-1)}\right) \tag{5}$$

2) END USER'S SERVICE UTILITY

In a traditional wireless network, to evaluate the service utility for different service requests, a Sigmoid function [30] based on the average transmission rate of service i. is represented as:

$$U\left[\bar{R}_{(i)}(t)\right] = \frac{1}{1+e^{-a\left[\bar{R}_{(i)}(t)-b\right]}} \tag{6}$$

where, a represents the slope of the Sigmoid utility function. b represents the gradient midpoint of the Sigmoid utility function, and $\bar{R}_{(i)}(t)$ represents the average transmission rate of service i at time slice t. The curves of the Sigmoid utility

Algorithm 1 Credit Updating Algorithm

Initialize the end user's credits $C_{E(k)}(t)$=0;
Initialize the control entity's credits $C_C(t)$=$T_C(t)$;
The control entity has sufficient accessing service time;
if $T_C(t) \geq \sum_{k=1}^{K} T_{R(k)}(t)$ **then**
  for $k = 1; k \leq K; k++$ **do**
    The control entity allocates credits to end user k:
    $C_{A(k)}(t) = T_{R(k)}(t)$;
    End user k obtains the accessing service time:;
    $T_{A(k)}(t) = T_{R(k)}(t)$;
    End user k accumulates credits:
    $C_{E(k)}(t) = C_{E(k)}(t-1) + C_{A(k)}(t)$;
    The control entity updates its the accessing service time:
    $T_C(t) = T_C(t) - \sum_{k=1}^{K} T_{R(k)}(t)$;
The control entity does not have sufficient accessing service time;
if $T_C(t) < \sum_{k=1}^{K} T_{R(k)}(t)$ **then**
  for $k = 1; k \leq K; k++$ **do**
    The control entity allocates credits to end user k: $C_{A(k)}(t) = C_C(t) \cdot \frac{F_{SI(k,i)}(t)}{\sum_{k=1}^{K} F_{SI(k,i)}(t)}$;
    End user k accumulates credits: $C_{E(k)}(t) = C_{E(k)}(t-1) + C_{A(k)}(t)$;
    End user k obtains the accessing service time: $T_{A(k)}(t) = T_C(t) \cdot \frac{C_{E(k)}(t)}{\sum_{k=1}^{K} C_{E(k)}(t)}$;
    The control entity updates its accessing service time:$T_C(t) = 0$;
return The control entity updates its the accessing service time;

function for data service, voice service, and streaming media service are shown in Fig.4.

However, in an IoV system, there are more complicated factors influencing the service scheduling decision, not just $\bar{R}_{(i)}(t)$. Each time an end user obtains the opportunity for network access, its service requirements can be updated dynamically based on its historical transmissions. Based on the Sigmoid utility function and fairness index $F_{CI(k,i)}(t)$ in Eq(5), we propose a new utility estimation function $U_{E(k,i)}(t)$ for end user k requesting service i as follows:

$$U_{E(k,i)}(t) = \frac{e^{-a[F_{CI(k,i)}(t)-b]}}{\left\{1 + e^{-a[F_{CI(k,i)}(t)-b]}\right\}^2} \tag{7}$$

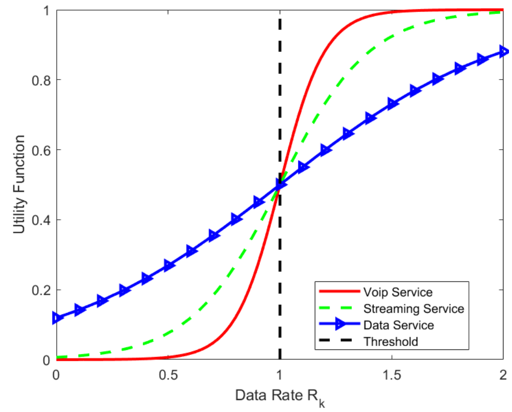


FIGURE 4. The curves of the Sigmoid utility function.

where, a represents the degree of sensitivity of $U_{E(k,i)}(t)$ to $F_{CI(k,i)}(t)$ and b represents the function inflection, which is the maximum degree of tolerance of service i to $F_{CI(k,i)}(t)$. The influence of a and b on $U_{E(k,i)}(t)$ is shown in Fig.5.

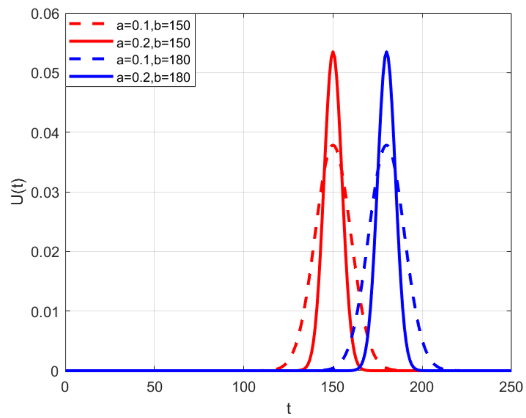


FIGURE 5. The curves of the utility estimation function.

V. NETWORK CODING SET AND CODING CACHE QUEUE

In this paper, we combine network coding with fair service resource scheduling. The challenge is that the basic unit of service scheduling is no longer a single end user, but a network coding set consisting of multiple end users. The solution for the above challenge can be modeled as a member selection method for the network coding set and corresponding optimization algorithm for the coding cache queue.

First, in the layer for the network coding set, we establish multiple network coding sets as the basic units of service scheduling, calculate the service utility for each network coding set, and perform network coding in each set.

Second, in the layer for the control entity, we construct a coding cache queue in the control entity based on the coding space information fed back by the member end users.

The following definitions are related to the network coding method used in this paper:

Definition 1 (Original Data Packet): And original data packet comes from an end user that has not obtained the opportunity to take part in the procedure of network coding. The original data packet is used to generate a coded data packet.

Definition 2 (Coding Space): The coding space is the vector space consisting of coding coefficient vectors in the finite field, and represents the decoding capabilities of the end user and control entity.

Definition 3 (Coding Cache Queue): The coding cache queue is maintained in the control entity for caching the original data packets during the exchange of coded data packets. The cache cost should be constrained because the storage of the control entity is limited.

A. MEMBER SELECTION FOR THE NETWORK CODING SET (LAYER FOR THE NETWORK CODING SET)

The principle of the member selection method of the network coding set is that: the member end users of a network coding set have similar attributes for credit and fairness index. **First**, all the end users are selected into multiple network coding sets by the member selection method. **Second**, we calculate the service utility for the network coding set based on the utility of the end user and the weight of the service attribute. **Third**, each network coding set generates coded data packets and constructs a coding space.

1) MEMBER SELECTION METHOD

It is assumed that all vehicles (end users) with a high velocity can maintain the connection with MEC servers (control entities) through the wireless channel of the RSU at all times. The communication range of the RSUs covers the whole road, as shown in Fig.6. Therefore, we can ensure that each end user belongs to a network coding set through the member selection method. The process of the member selection method is divided into four steps:

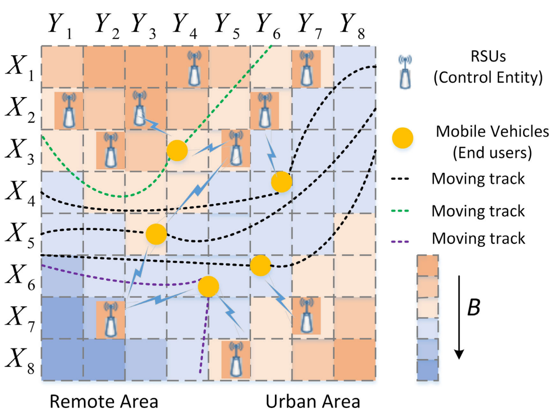


FIGURE 6. Network model for communication between end users and control entities.

Step 1: Set a member threshold for the network coding set

We set a threshold G_X for the members of a network coding set to achieve a balance between the decoding efficiency and the space of the coding cache queue.

Step 2: Sort and select the end users

During member selection, all the end users are sorted into two lists in descending order according to their credits and the fairness index. For the above two lists, we set an interval with a length of G_X and search for the end users that are simultaneously in these two lists. The common end users are added to one network coding set until the number of common users reaches G_X.

Step 3: Construct multiple network coding sets

If the number of common users in the above two lists exceeds G_X, a new network coding set is established. Then we perform network coding in the network coding set. Coded data packets are generated based on the original data packets from the member end users.

Step 4: Update the composition of the network coding set

After transmitting the coded data packets, each end user's credits and fairness index change accordingly. Go to Step 2, and new network coding sets are established.

2) UTILITY OF NETWORK CODING SET

An end user may request multiple services. For each service, we consider its multi-dimensional attributes, including transmission cost, response time, throughput, availability, reliability, etc. [31]. Each attribute has a weight ω_1. At time slice t, $U_{(k)}(t)$ is the utility for end user k based on Eq(7), which is represented as:

$$U_{(k)}(t) = \sum_{i=1}^{I} \omega_l \cdot U_{E(k,i)}(t) \tag{8}$$

Then, for a network coding set consisting of G_X member end users, the utility of the network coding set is defined as $U_S(t)$:

$$U_S(t) = \sum_{k=1}^{G_X} \left\{ U_{(k)}(t) \right\} \tag{9}$$

3) NETWORK CODING MECHANISM

We adopt a classical inter-session wireless network coding scheme referred as COPE [32], the classic random linear network coding method in this paper. Assume that the IoV system ensures that the coding space of each control entity is large enough, and that all the original data packets can be recovered.

The network coding mechanism is as follows: **First**, the control entity calculates its current coding space according to the feedback from the end users. **Second**, according to the threshold G_T, multiple network coding sets are established by member selection, and all the end users will be the members of different network coding sets. **Third**, each network coding set generates and exchanges coded data packets with each other, and the destination end users recover the original data packets based on the coding space's basis matrix of the control entity.

For the network coding mechanism, the transformation process of the coding space's basis matrix in the network coding set is as follows:

Step 1: Initialize the original coding space.

K empty matrices $MT_1, \ldots, MT_K$ in the control entity represent K end users. The information of T new original data packets is added to the matrices maintained by the control entity. $(K - T)$ elements of 0 are added to each row of matrix MT_k $(k = 1, \ldots, K)$ accordingly to complete the initialization of the original coding space.

Step 2: Generate coded data packet and exchange the coding space information.

Using the heuristic network coding method COPE, a coded data packet $P_{S(m)}$ is generated based on the original data packets from the member end users, and $P_{S(m)}$ is exchanged among network coding sets.

Step 3: Transmit the coding space information to control entity.

First, if end user k $(k = 1, \ldots, K)$ in the coding set correctly receives a coded data packet in Step 2), the coding coefficient vector of the coded data packet is transferred to a row of matrix MT_k;

Second, conduct Gaussian elimination to obtain the basis matrix of the coding space for end user k;

Third, end user k transmits its coding space information to the control entity. Each network coding set can only transmit coded data packets while being allocated accessing service time.

Step 4: Update the basis matrix of the coding space.

Once the coding space information is received from the end users, the control entity updates the coding space MT_k. The information of the original data packets preserved by the end users is also updated, and a new basis matrix of the coding space is established; return to Step 2).

B. CODING CACHE QUEUE CONTROL (LAYER FOR THE CONTROL ENTITY)

In the layer for the control entity, the coding queue should provide cache service fairly for all the end users taking part in network coding. However, the control entity has limited cache space, and the cache queue of the control entity needs to pay a higher cost for caching the original data packets without reasonable queue control.

1) OPTIMIZATION OF THE UTILITY OF THE CONTROL ENTITY

A network coding set is the basic unit of service scheduling. The optimization objective of the scheduling service is to maximize the total allocation utility of all the network coding sets (the utility of the control entity), then the control entity can make the scheduling decision and construct the coding cache queue accordingly. The utility of the control entity U_C based on Eq(9) is defined as:

$$U_C = \sum_{m=1}^{M} \left\{ U_S \left[S_{(m)} \right] \right\} \tag{10}$$

where, M is the number of network coding sets in the control entity.

First, we define a six-tuple $G = \{E, V, A, X, P, W\}$ as the basic model of system resource scheduling as follows:

(1) End user set $E = \{e_{(k)} | e_{(1)}, e_{(2)}, \ldots, e_{(K)}\}$ with K end users.

(2) Service set $V = \{v_{(i)} | v_{(1)}, v_{(2)}, \ldots, v_{(I)}\}$ with I types of services.

(3) Service request set $A = \{a_{(k,i)} | a_{(1,1)}, a_{(1,2)}, \ldots, a_{(K,I)}\}$ represents that end user k requests service i. $a_{(k,i)} = 0$ represents that end user k does not request service i.

(4) Task time set $X = \{x_{(k,i)} | x_{(1,1)}, x_{(1,2)}, \ldots, x_{(K,I)}\}$ represents the execution time of each service request $a_{(k,i)}$.

(5) Service cost set $P = \{p_{(i)} | p_{(1)}, p_{(2)}, \ldots, p_{(I)}\}$ for each service i.

(6) Service weight set $W = \{\omega_i | \omega_1, \omega_2, \ldots, \omega_I\}$.

Second, we construct a scheduling decision matrix for all the end users. $r_{(k,i)} = a_{(k,i)} \cdot x_{(k,i)} \cdot v_{(i)}$ represents the service budget for each end user k with service i, so the scheduling decision matrix for all the end users is:

$$\begin{bmatrix} r_{(1,1)} & r_{(1,2)} & \cdots & r_{(1,I)} \\ r_{(2,1)} & r_{(2,2)} & \cdots & r_{(2,I)} \\ \vdots & \vdots & \ldots & \vdots \\ r_{(K,1)} & r_{(K,2)} & \cdots & r_{(K,I)} \end{bmatrix} \tag{11}$$

Then we normalize the scheduling decision matrix as follows:

$$r_{N(k,i)} = \frac{\max r_{(k,i)} - r_{(k,i)}}{\max r_{(k,i)} - \min r_{(k,i)}} \tag{12}$$

where $r_{N(k,i)}$ is the normalized form of the service budget $r_{(k,i)} = a_{(k,i)} \cdot x_{(k,i)} \cdot p_{(i)}$.

Third, the optimization objective of the scheduling service is to maximize the utility of the control entity with a limited service budget and a high requirement for service execution time from the end users, and this objective is represented as:

$$\begin{aligned} \max U_C &= \max \sum_{m=1}^{M} \left\{ U_{(k)} \left[S_{(m)} \right] \right\} \\ &= \max \sum_{m=1}^{M} \sum_{k=1}^{G_X} U_{(k,i)} \left[S_{(m)} \right] \\ &= \max \sum_{m=1}^{M} \sum_{k=1}^{G_X} \sum_{i=1}^{I} \omega_i \cdot F_{I(k,i)}(t) \end{aligned}$$

$$s.t. \begin{cases} \sum_{i=1}^{I} [a_{(k,i)} \cdot x_{(k,i)} \cdot v_{(i)}] \le C_{(k)} \\ \sum_{i=1}^{I} [a_{(k,i)} \cdot x_{(k,i)}] \le D_{(k)} \\ 0 \le x_{(k,i)} \le 1, \sum_{i=1}^{I} x_{(k,i)} = 1 \end{cases} \tag{13}$$

where, $C_{(k)}$ represents the maximum service budget for end user k from the system and $D_{(k)}$ represents the requirement for service execution time from end user k.

The values of the weights of the service attributes ω_l are important for maximizing the utility of the control entity and the following coding cache queue control. Based on the above definitions, the optimization process for the utility of the control entity based on the Lagrange relaxation factor [33] is divided into 5 steps as follows:

Step 1: Initialize the six-tuple G, the maximum budget vector $C = \{C_{(1)}, C_{(2)}, \ldots, C_{(K)}\}$, and the constrained vector $D = \{D_{(1)}, D_{(2)}, \ldots, D_{(K)}\}$ for the service finishing time.

Step 2: The end user transmits the service requirement information to the control entity. The control entity evaluates the service capability and constructs the scheduling decision matrix.

Step 3: The control entity calculates its utility function with Eq(13), and introduces Lagrange relaxation factor λ to relax the constraint conditions to obtain an optimization function without constraints.

Step 4: The control entity obtains the service weight set $W = \{\omega_i | \omega_1, \omega_2, \ldots, \omega_I\}$, and makes the system resource scheduling decision.

Step 5: Repeat Step 2 to Step 4 until all the end users obtain system service resources from the control entity.

2) CODING CACHE QUEUE CONTROL WITH QPSO

End users with various service attributes have different cache requirements for performing network coding. After determining the weights of the service attributes $W = \{\omega_1, \omega_2, \ldots, \omega_I\}$ for maximizing the utility, the control entity constructs a coding cache queue.

The challenge of the classic Proportional Integral (PI) model for active queue control [34] is to adaptively set two key parameters: a proportional coefficient k_P and a integral coefficient k_I. k_P can change the system state quickly, and k_I can reduce the system static errors. We introduce a Quantum Particle Swarm Optimization (QPSO) algorithm [35] to obtain the global optimal solution for k_P and k_I, and control the queue's length for the control entity.

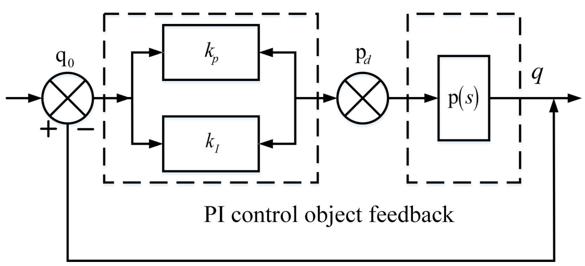


FIGURE 7. Proportional Integral (PI) model.

The PI model is shown in Fig.7. At time slice t, q_E is the queue's expected length. $q_L(t)$ is the queue's maximum length. $E_d(t)$ is the variable deviation. $q_D(t)$ is the packet dropping rate, and $\Delta q_D(t)$ is the increment of the packet dropping rate; these variables are shown as [34]:

$$\begin{cases} E_d(t) = q_E - q_L(t) \\ q_D(t) = k_P \cdot E_d(t) + q_L \cdot \sum_{j=1}^{t} E_d(j) \\ \Delta q_D(t) = k_P \cdot \left[\left(1 + T_S \cdot \frac{k_I}{k_P}\right) \cdot E_d(t) - E_d(t-1) \right] \end{cases} \tag{14}$$

The process of solving the optimal values for the PI model using the QPSO algorithm is divided into 5 steps as follows:

Step 1: Initialize the parameters of QPSO, including the number of particle swarms S, the population size of a particle swarm H, the maximum number of iteration $k_{\max}$, and the current sampling time.

Step 2: Generate particle swarm $J = \{\mathbf{j}_{(s)} | \mathbf{j}_{(1)}, \mathbf{j}_{(2)}, \ldots, \mathbf{j}_{(S)}\}$ consisting of SH particles, where each particle is a two-dimensional vector $\mathbf{j}_{(s)} = \{k_{P(s)}, k_{I(s)}\}$. Search for the optimal value of $\mathbf{j}_{(s)}$ based on the QPSO algorithm, and reset the PI model.

Step 3: Calculate each particle's fitness based on its performance index.

Step 4: Sample the current queue's length $q_L(t)$, and calculate $\Delta q_D(t)$ based on Eq(14). Calculate the optimal position for each single particle [35].

Step 5: Update the iteration number. Output k_P and k_I when the iteration number reaches $k_{\max}$; other wise return to Step 1.

VI. THE EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we conduct simulations to demonstrate the performance of the proposed DRSO algorithm and the general solution framework for fair service scheduling in the IoV system. We compare DRSO with the following approaches: minimum-latency aggregation scheduling algorithm (MLAS) [36], balanced shortest path tree (BSPT) for efficient scheduling [37], round robin pairing and scheduling algorithm (RRPS) [38], distributed data aggregation scheduling algorithm (DDAS) [39] and fast data aggregation protocol (FDAP) for collision-free schedule [40]. Among these approaches, MLAS employs connected dominating sets and maximal independent sets (DRSO selects end users into multiple network coding sets). BSPT constructs a logical tree allowing the generation of short schedules for the top-k queues (DRSO establishes a coding queue based on the scheduling decision). RRPS selects a pair of users carrying out transmission simultaneously in uplink virtual multiple input multiple output (DRSO selects the member end users of a network coding set with similar attributes for credits and fairness index). DDAS generates a collision-free schedule for data aggregation (DRSO constructs a network coding set as the basic unit of service scheduling). FDAP proposes criteria for node selection among available competitors to reduce the time latency of the aggregation schedule (DRSO sets a threshold for the members of a network coding set to achieve a balance between the decoding efficiency and the space of the coding cache queue).

We conduct tests on different benchmarks of IoV services. The purpose of this study is to investigate the following performance indexes: (1) fairness index; (2) network throughput; (3) packet loss rate; (4) time slice allocation; (5) service scheduling rate; (6) average system response time; (7) average end-to-end delay.

A. SIMULATION SETTINGS

Network simulator 2 (NS2) is used to conduct extensive simulations. The simulation environment is set up based on the system architecture described in Section III. The detailed simulation parameters are listed in Table. 3.

TABLE 3. Simulation parameters.

Parameters	Description
Packet size	2k Bytes
Average data transmission rate	2 Mbps
Time slice	1ms
Transmission power	30 dBm
Noise power	-50 dBm
SNR threshold	15 dB
Maximum queue length	200 packets
Maximum cache capacity	200 packets
Workload	7 gigacycles
Simulation time	200s, 500s

To thoroughly evaluate the performance of DRSO, a public vehicle driving dataset containing GPS trajectories from amap.com is adopted in our experiments [41]. A movement trace map of vehicles and a traffic heat-map from Mar. 24 to Mar. 26, 2020 within Beijing are given in Fig.8 and Fig.9.

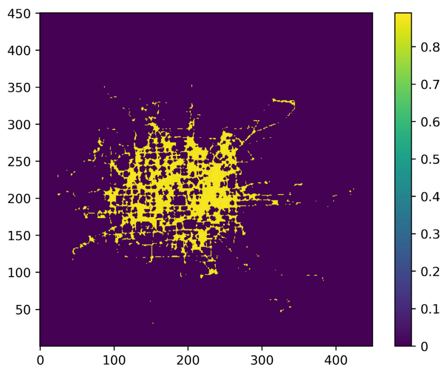


FIGURE 8. Visualization of traffic trajectory dataset of Beijing.

There are three requirements for the driving trajectory data used in the experiments: **First**, the acceleration satisfies a normal distribution. **Second**, the driving state of vehicles is stable most of the time. **Third**, there are sufficient end users (vehicles) to form network coding sets with the threshold $G_X = 40$ in a multi-user IoV system [42]. To satisfy the above requirements, we extract the records in a period of time according to the trajectory data of Beijing, and obtain the corresponding histograms of the relative probability distribution for the acceleration and the number of end users (vehicles).

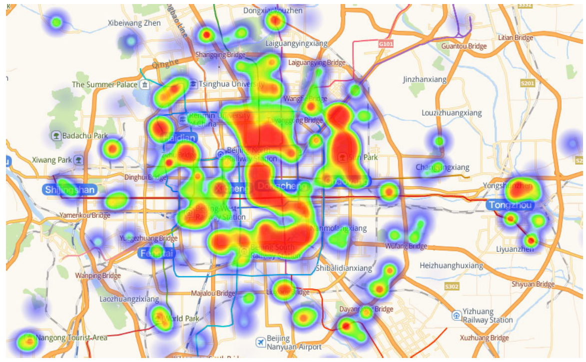


FIGURE 9. Traffic heat map of Beijing.

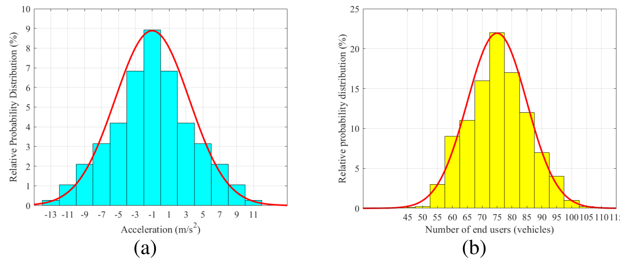


FIGURE 10. Relative probability distribution for acceleration and the number of end users (vehicles).

The RSU (the control entity) is 50 m away from the center of the nearest lane. The maximum communication range of an RSU and an end user (vehicle) is 200 m. Fig. 10(a) shows that the acceleration presents an approximately normal distribution, the values of the acceleration is distributed mainly in the range between $7m/s^2$ and $-7m/s^2$ and the vehicles seldom have rapid acceleration and deceleration. Fig. 10(b) shows that the number of end users (vehicles) is mainly distributed between 55 to 100 on the two-way lane within the limited communication range. Most vehicles have a stable driving state, which can help us study driving state. In summary, we can use the trajectory data of Beijing for the following simulations.

It is assumed that all the end users are synchronized in both time slices and transmission frames [43]. The IEEE 802.11p standard is employed to support the architecture of the multi-user IoV system, with a bandwidth of 10 MHz per channel. The vehicles send data packets at the lowest rate of 3 Mbps, as this provides the best service reliability [44]. The large-scale path loss is characterized by a Nakagami fading factor model of the IoV channel. The main service types used in this paper include: real-time Streaming Media Service (**SMS**), VoIP Session Service (**VSS**) and non-real-time Best Effort Service (**BES**).

B. SIMULATION RESULT ANALYSIS

1) FAIRNESS INDEX

Now we evaluate the impact of the number of end users on the fairness index for SMS, VSS and BES in Fig. 11.

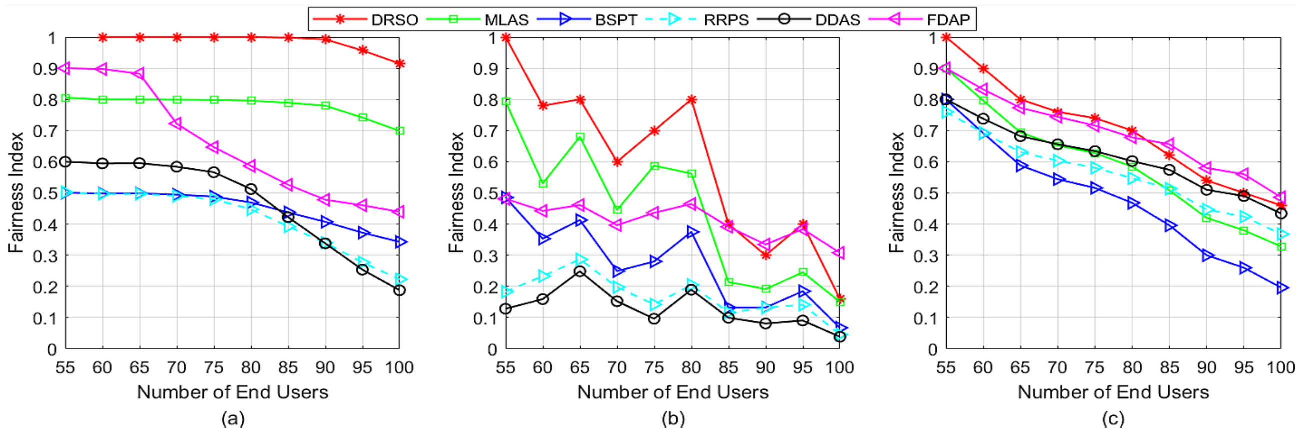


FIGURE 11. Fairness index with different numbers of end users. (a) SMS. (b) VSS. (c) BES.

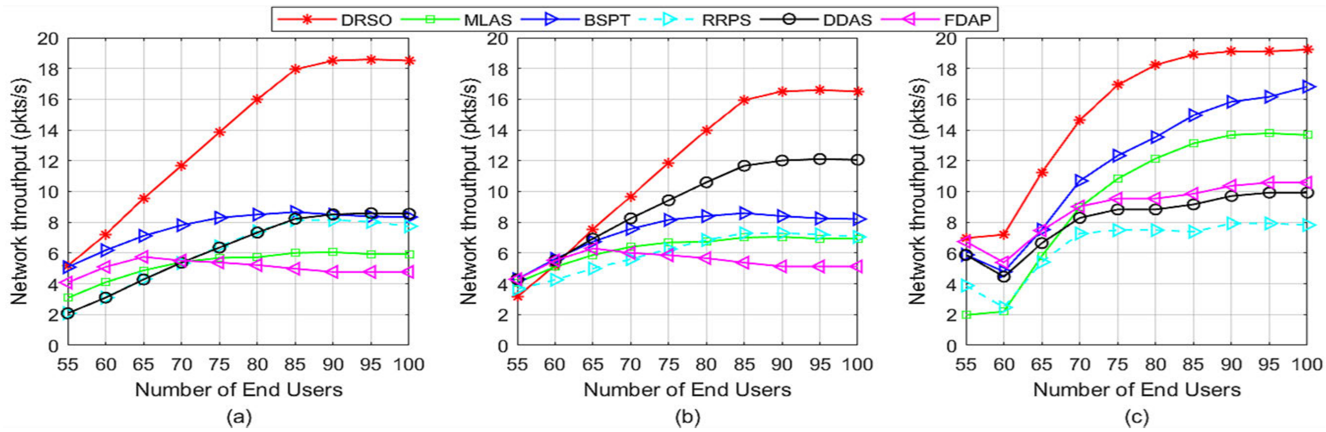


FIGURE 12. Network throughput with different numbers of end users. (a) SMS. (b) VSS. (c) BES.

For SMS, Fig. 11(a) shows that, when the number of end users is less than 60, the fairness index of most algorithms can be maintained at a high level. This is because all the service requests can be satisfied when the control entity has sufficient accessing service time. However, as the number of end users increases, the fairness index decreases. This occurs because as more end users compete for system resources, the accessing service time provided by the control entity becomes insufficient. Affected by both the number of end users and the completion time of a service request, some end users may lose the opportunity to access the system. DRSO introduces a compensation factor C_F in Eq(2) that gives a high service priority to end users with strict SMS requirements. Thus, DRSO can ensure that all the end users requesting SMS have a high fairness index.

In terms of VSS, Fig.11(b) shows that the fairness index of each algorithm suffers transmission fluctuations. This is because VSS is not a persistent service. DRSO can maintain better performance than the other algorithms because DRSO adopts the statistical fairness index $F_{SI(k,i)}(t)$ in Eq(4), which can reduce the negative impact caused by the short-term fluctuation of wireless communication in the IoV. Since the other algorithms do not optimize the service scheduling order, it is hard for end users with higher delay costs to obtain the system resources.

As shown in Fig.11(c), BES is a low-priority service, and usually experiences congestion as the number of end users increases. The fairness index of all algorithms decreases with the number of vehicles. DRSO still obtains a higher fairness index because the network coding in DRSO can reduce the retransmission times of data packets. Thus, DRSO can help most end users reduce their service completion time.

It is observed that in general, a user's fairness index decreases along with an increasing number of end users. This reveals that scheduling-based algorithms perform better than random access methods because the chance of obtaining channel access is influenced by the distance between end users in random access algorithms especially in an IoV system. In DRSO based on fair scheduling, since the accessing service time is ordered and scheduled regardless of the end user's location, good fairness performance can be obtained.

2) NETWORK THROUGHPUT

As shown in Fig.12, the network throughput is investigated with the number of end users. It is observed that in general

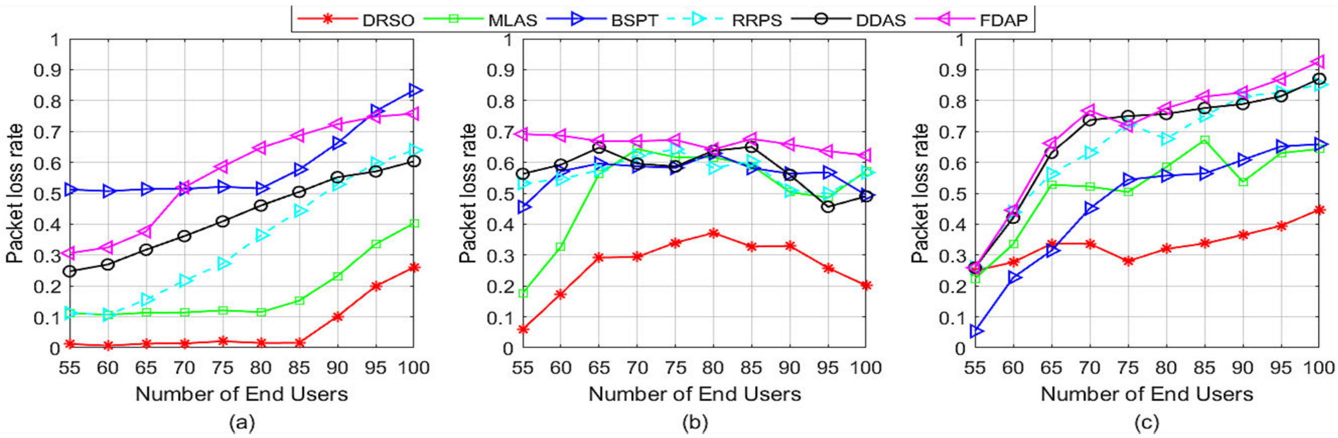


FIGURE 13. Packet loss rate with different numbers of end users. (a) SMS. (b) VSS. (c) BES.

the throughput increases along with the number of end users, and the throughput of all the algorithms has an upper bound.

As shown in Fig.12(a), when the number of end users reaches 85, each algorithm has nearly reached its upper bound. The throughput of the proposed DRSO algorithm is better than that of the other algorithms. More specifically, DRSO performs better than BSPT, indicating that fair scheduling-based transmission with network coding outperforms random scheduling methods. Moreover, the performance of DRSO is much better than that of the other scheduling algorithms, suggesting that the introduction of network coding is useful in this network model.

In terms of Fig.12(b), all the algorithms exhibit similar throughput performance trends. More end users lead to higher workloads in the system. The throughput of each algorithm increases because higher workloads lead to the generation of more data packets for VSS.

Fig.12(c) shows that, the throughput of BES with low service priority exhibits a decline for all the algorithms when there are not many end users, because the chances of accessing the system are reduced. However, the throughput of DRSO is still superior to that of the other algorithms, especially under circumstance with heavy offered traffic loads. The network coding occurrence of DRSO is higher than that of the other algorithms. The coding cache queue in DRSO processes the coded data packets more quickly, accommodates more coded data packets for exchanging and postpones the network congestion, and these factors lead to the throughput advantage of DRSO compared with the other algorithms.

3) PACKET LOSS RATE

Fig.13 shows the changes in the packet loss rates of the algorithms with the number of end users.

As observed in Fig.13(a) for SMS, the packet loss rate increases when the traffic workload starts to become heavier to a certain extent. As the number of end users increases, the collision probability among data packets rises and the packet loss rates of all the algorithms continue to increase. The reasons are explained as follows. At the beginning, all algorithms can achieve a low packet loss rate due to the small amount of data flows. When the service requests from end users starts start to increase, the increased number of packet collisions dominate the performance, which results in an increase in the packet loss rate.

As shown in Fig.13(b) for VSS, the packet loss rate of DRSO is lower than that of the other algorithms as the number of end users increases. This is because fair scheduling gradually plays a useful role in avoiding packet collision and network coding can efficiently reduce retransmission. This demonstrates the advantages of network coding and the fair scheduling strategy adopted by DRSO.

As shown in Fig.13(c) for BES, the packet loss rate of DRSO is lower than that of the other algorithms in most of cases, while it is higher than BSPT at the beginning. This is because the advantages of stable scheduling are not obvious in the case of a small amount of packet transmission. DRSO exhibits fairness adaptive control of data packet transmission. When the number of end users reaches 65, stable scheduling plays an important role, and the packet loss rate of DRSO is better than that of BSPT.

4) TIME SLICE ALLOCATION FOR ACCESSING THE SYSTEM

Fig.14 presents the time slice allocation of all the algorithms under different channel loads. We compare the time slices of service requests and the time slices allocated for the three types of services (SMS, VSS and BES) for the different algorithms. When the channel load is low, all the service requests for time slices can be satisfied. However, the total accessing service time provided by the control entity is constrained. Therefore, the service request for time slices cannot be fully satisfied with an increase in channel load.

Fig.14(a) shows the time slices provided by the control entity with DRSO. SMS with its high service priority can obtain more time slices than VSS and BES, especially in the case of a heavy channel load. The differences in the

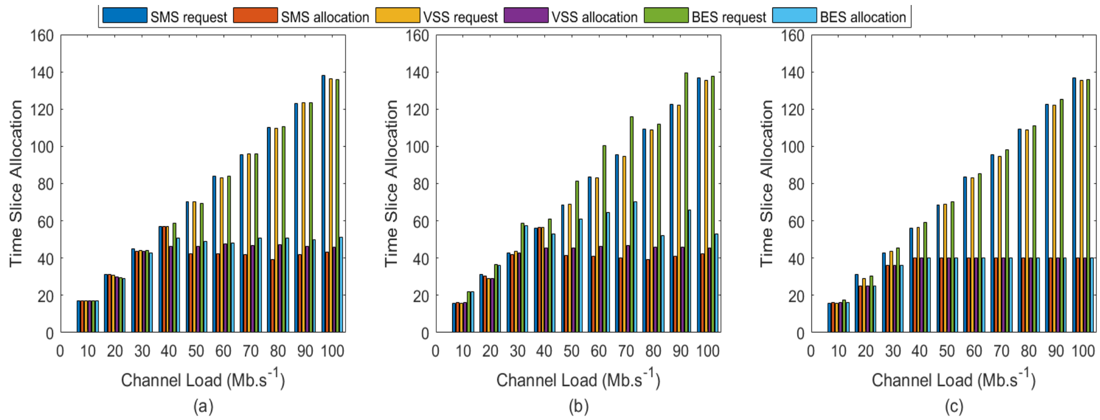


FIGURE 14. Time slice allocation for accessing system for different services. (a) DRSO. (b) BSPT. (c) DDAS.

time slices allocated for the three types of services can be controlled in DRSO, even as the channel load increases. Fig.14(b) shows the time slices provided by the with BSPT. The difference in the time slices allocated for the three types of services in BSPT are more significant than those in DRSO. BSPT does not consider fairness among end users with different services. Fig.14(c) presents the time slices provided by the control entity with DDAS, and there is no difference among the time slices allocated for three types of services. Therefore, the different service requirements cannot be fully satisfied.

It is observed that DRSO outperforms the other algorithms. The reason behind this phenomenon is that, DRSO can detect more coding opportunities because all the end users are selected into network coding sets, as described in Section V, and obtain a fair opportunity to take part in the activity of network coding. Furthermore, DRSO introduces a credit updating algorithm in the layer for the end user in Section IV. Credit can be used to balance service resources among the end users.

5) SERVICE SCHEDULING RATE FOR END USERS

The average scheduling rates for end users with different types of services (SMS, VSS and BES) in DRSO and BSPT are shown in Fig.15. The scheduling objects of DRSO and BSPT both satisfy the expected scheduling rates of the end users. End users differ in channel state and service scheduling requirements. The IoV system is in a state of heavy channel load when the sum of the expected service scheduling rates of all the end users reaches 90% of the overall downlink throughput of the IoV system. Therefore, we define the Satisfaction Degree (SD), which is the ratio of the cumulative average scheduling rate to the expected scheduling rate.

As shown in Fig.15(a), the SD of most end users reached 98%, and most end users obtained the expected service scheduling rate. However, as observed in Fig.15(b), the SD of most end users only reaches 85%. Therefore, DRSO is superior to BSPT in satisfying the expected service scheduling rate of each end user in a heavy channel load scenario. DRSO is more efficient in using channel resources.

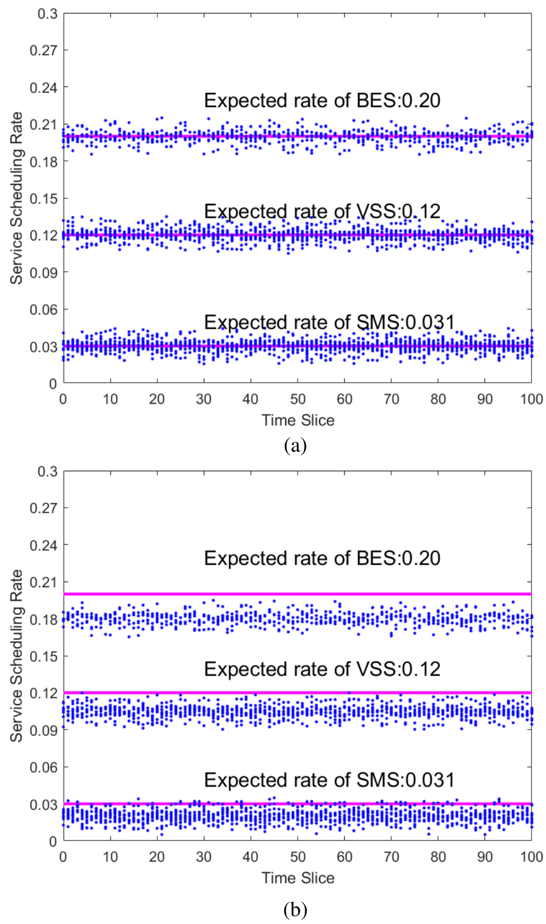


FIGURE 15. Average scheduling rates for different types of services. (a) DRSO. (b) BSPT.

The reason is as follows: as the channel load rises, more data packets can intersect with each other to generate coded data packets, and more coded data packets can be scheduled. DRSO transmits the coded data packets by stable scheduling and fair allocation of available resources for accessing service time. BSPT ignores the problem of stable scheduling, and the variability of service requirements from end users, so most service requirements cannot be fully satisfied. Unlike DRSO, BSPT is not always able to transmit the data packets in a uniform state, since it does not explore the fair allocation of available accessing service time.

6) AVERAGE SYSTEM RESPONSE TIME

The Average System Response Time (ASRT) is one of the main indicators for evaluating scheduling algorithms. In a dynamic IoV environment, a random channel causes instability in the ASRT. We treat the ASRT as a random variable, and we also expect the variance of the ASRT to be as small as possible. Therefore, this variance is also an important evaluation index in dynamic scheduling.

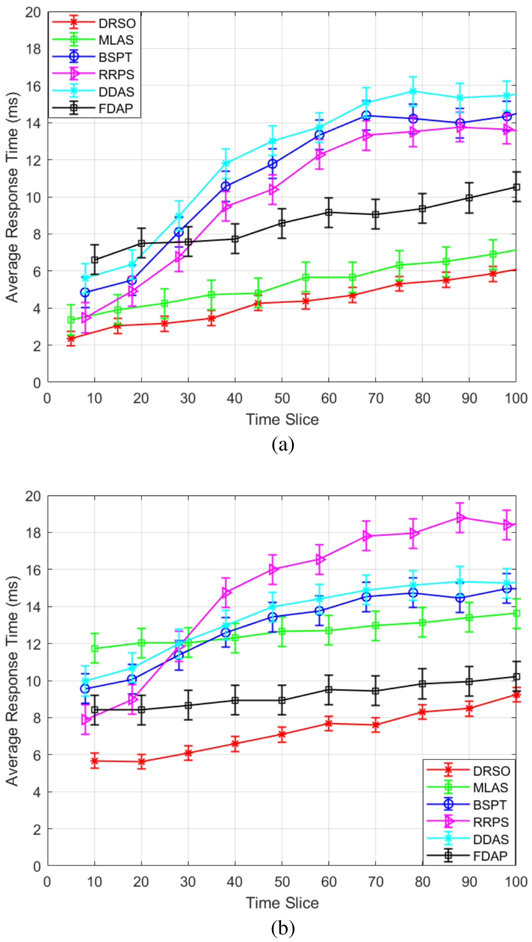


FIGURE 16. Average system response time with different time intervals of service request arrivals. (a) exponential distribution. (b) random distribution.

In Fig.16(a), it is assumed that the time interval of service request arrivals satisfies an exponential distribution, and the service requests have the same size. In this condition, DRSO achieves the best performance. The ASRT of DRSO is lower than that of the other algorithms. Since DRSO can find more coding opportunities through network coding sets, DRSO constructs the coding cache queue of coded data packets more quickly by more occurrence of network coding, which greatly reduces the number of retransmissions and the ASRT. Therefore, DRSO can accommodate more end users taking part in network coding, which promotes the efficiency of service scheduling.

In Fig.16(b), it is assumed that the time interval of service request arrivals satisfies a random distribution, and the service requests are not the same size. The ASRT of all the algorithms increases. DRSO can still keep its ASRT at a low level. The Other algorithms neglect fair scheduling during packet transmission, and end users requesting service at low priority may never obtain the chance for network coding, which leads to a rise in the ASRT.

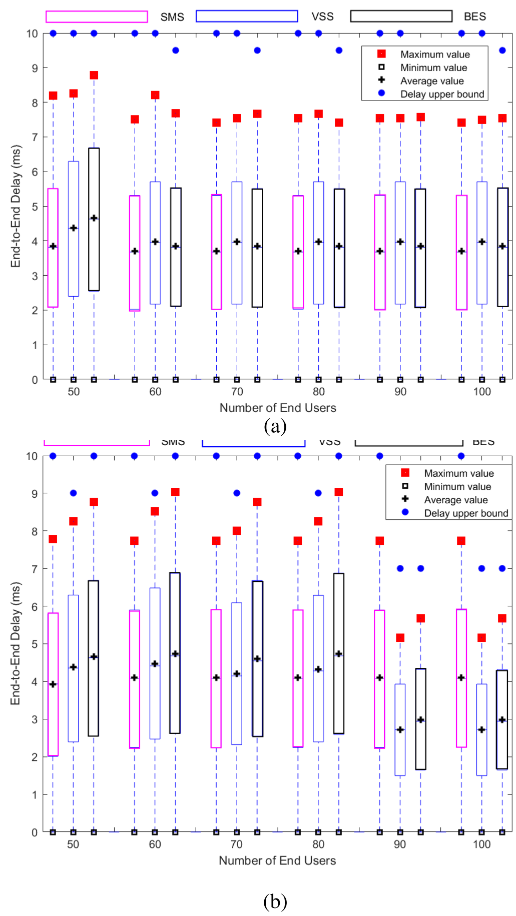


FIGURE 17. Average system response time with different time interval of service request arrivals. (a) BSPT. (b) DRSO.

7) AVERAGE END-TO-END DELAY

Fig.17 displays the average end-to-end delay of the three types of services (BES, VSS, and SMS) separately with DRSO and BSPT. We vary the number of end users from 50 to 100.

Fig.17(a) shows that, BSPT focuses only on the fairness of SMS with a high service priority. Most of the time, only end users requesting SMS can obtain a low end-to-end delay. As shown in Fig.17(b), in general, all three services can fairly obtain a low end-to-end delay with DRSO. DRSO performs better than BSPT. The reason is as follows: In DRSO, end users do not need to wait for accessing service time until being scheduled, and each member end user in the network coding set can fairly obtain accessing service time once the network coding set is scheduled by the control entity.

VII. CONCLUSION AND FUTURE WORK

In this paper, to achieve an adaptive balance between fairness and throughput in a multi-user IoV system, we have designed DRSO, a dynamic fair scheduling algorithm combined with network coding for system resource allocation. We construct a general solution framework with four separate layers for the service, end user, network coding set, and control entity, and we define a utility model for each of them. Then we formulate the utility maximization of the control entity (MEC servers) as the optimization objective of service scheduling, and make the optimal scheduling decision accordingly. Simulation results prove the good performance of DRSO. DRSO effectively improves the fairness, network throughput, packet loss rate, service scheduling rate, system response time, and end-to-end delay with varying traffic loads, and these findings demonstrate that DRSO can be used for guiding service resource allocation in multi-user IoV systems.

In the future, we would implement the proposed DRSO in a large scale with the cooperation among the end users, and we would also study individual end user response patterns to scheduling decisions, and analyze the correlation between system performance and these patterns. By doing so, we wish to obtain insights into the imbalance between service requests and resource allocation from a new perspective.

REFERENCES

[1] A. Naeem, M. H. Rehmani, Y. Saleem, I. Rashid, and N. Crespi, "Network coding in cognitive radio networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1945–1973, Jan. 2017.

[2] A. Douik, S. Sorour, T. Y. Al-Naffouri, and M. S. Alouini, "Delay reduction for instantly decodable network coding in persistent channels with feedback imperfections," in *IEEE Commun. Surveys Tutor.*, vol. 20, no. 2, pp. 1014–1035, Jan. 2018.

[3] H. Holbrook, S. Singhal, and D. Cheriton, "Log-based receiver-reliable multicast for distributed interactive simulation," *ACM SIGCOMM Comp. Commun. Rev.*, vol. 25, no. 5, pp. 328–341, Oct. 1995.

[4] P. Chaporkar and A. Proutiere, "Adaptive network coding and scheduling for maximizing throughput in wireless networks," in *Proc. 13th ACM Conf. Mobile Comput. Netw.*, 2007, pp. 135–146.

[5] C. Zhang, X. Liang, Z. Wu, F. Wang, S. Zhang, Z. Zhang, and X. You, "On the low-complexity, hardware-friendly tridiagonal matrix inversion for correlated massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6272–6285, Jul. 2019.

[6] I. Chatzigeorgiou and A. Tassi, "Decoding delay performance of random linear network coding for broadcast," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7050–7060, Aug. 2017.

[7] Q. Ding and Y. Jing, "Outage probability analysis and resolution profile design for massive MIMO uplink with mixed-ADC," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6293–6306, Sep. 2018.

[8] Z. Li and J. Gui, "Energy-efficient resource allocation with hybrid TDMA–NOMA for cellular-enabled Machine-to-Machine communications," *IEEE Access*, vol. 7, pp. 105800–105815, 2019.

[9] S. Shahsavari, F. Shirani, and E. Erkip, "A general framework for temporal fair user scheduling in NOMA systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 408–422, Jun. 2019.

[10] H.-W. Ferng, C.-Y. Lee, J.-J. Huang, and Y.-J. Liang, "Urgency-based fair scheduling for LTE to improve packet loss and fairness: Design and evaluation," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2825–2836, Mar. 2019.

[11] G. Liu, L. Li, L. J. Cimini, and C.-C. Shen, "Extending proportional fair scheduling to buffer-aided relay access networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 1041–1044, Jan. 2019.

[12] M. O. Ojo, S. Giordano, D. Adami, and M. Pagano, "Throughput maximizing and fair scheduling algorithms in industrial Internet of Things networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 6, pp. 3400–3410, Jun. 2019.

[13] K. Li, C. Yuen, B. Kusy, R. Jurdak, A. Ignjatovic, S. S. Kanhere, and S. Jha, "Fair scheduling for data collection in mobile sensor networks with energy harvesting," *IEEE Trans. Mobile Comput.*, vol. 18, no. 6, pp. 1274–1287, Jun. 2019.

[14] M. Ge and D. M. Blough, "High throughput and fair scheduling for multi-AP multiuser MIMO in dense wireless networks," *IEEE/ACM Trans. Netw.*, vol. 26, no. 5, pp. 2414–2427, Oct. 2018.

[15] F. Liu and M. Petrova, "Performance of proportional fair scheduling for downlink PD-NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 7027–7039, Oct. 2018.

[16] H. Li and X. Huang, "Multicast systems with fair scheduling in non-identically distributed fading channels," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8835–8844, Oct. 2017.

[17] Y. Chen, X. Wang, and L. Cai, "On achieving fair and throughput-optimal scheduling for TCP flows in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 7996–8008, Dec. 2016.

[18] Z. Hadzi-Velkov, I. Nikoloska, H. Chingoska, and N. Zlatanov, "Proportional fair scheduling in wireless networks with RF energy harvesting and processing cost," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2107–2110, Oct. 2016.

[19] D. Parruca and J. Gross, "Throughput analysis of proportional fair scheduling for sparse and ultra-dense interference-limited OFDMA/LTE networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6857–6870, Oct. 2016.

[20] J. Gu, S. J. Bae, S. F. Hasan, and M. Y. Chung, "Heuristic algorithm for proportional fair scheduling in D2D-cellular systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 769–780, Jan. 2016.

[21] M. Chen, T. Wang, K. Ota, M. Dong, M. Zhao, and A. Liu, "Intelligent resource allocation management for vehicles network: An A3C learning approach," *Comput. Commun.*, vol. 151, pp. 485–494, Feb. 2020.

[22] M. Huang, A. Liu, N. N. Xiong, T. Wang, and A. V. Vasilakos, "An effective service-oriented networking management architecture for 5G-enabled Internet of Things," *Comput. Netw.*, vol. 173, May 2020, Art. no. 107208.

[23] T. Li, M. Zhao, and K. K. L. Wong, "Machine learning based code dissemination by selection of reliability mobile vehicles in 5G networks," *Comput. Commun.*, vol. 152, pp. 109–118, Feb. 2020.

[24] J. Li, D. Su, and Y. Wang, "Energy-efficient and traffic-adaptive Z-medium access control protocol in wireless sensor networks," *IET Wireless Sensor Syst.*, vol. 8, no. 5, pp. 208–214, Oct. 2018.

[25] A. Salem, C. Masouros, and K.-K. Wong, "Sum rate and fairness analysis for the MU-MIMO downlink under PSK signalling: Interference suppression vs exploitation," *IEEE Trans. Commun.*, vol. 67, no. 9, pp. 6085–6098, Sep. 2019.

[26] M. Mohseni, S. A. Banani, A. W. Eckford, and R. S. Adve, "Scheduling for VoLTE: Resource allocation optimization and low-complexity algorithms," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1534–1547, Mar. 2019.

[27] M. Sefunc, A. Zappone, and E. A. Jorswieck, "Energy efficiency of mmWave MIMO systems with spatial modulation and hybrid beamforming," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 1, pp. 95–108, Mar. 2020.

[28] J.-Y. Huang and H.-F. Lu, "Achieving large sum rate and good fairness in MISO broadcast communication," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 5684–5695, Jun. 2019.

[29] T. Alshammari, B. Hamdaoui, M. Guizani, and A. Rayes, "Malicious-proof and fair credit-based resource allocation techniques for DSA systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 606–615, Feb. 2015.

[30] Z. Dai, P. Wang, H. Wei, and Y. Xu, "Adaptive detection with constant false alarm ratio in a non-Gaussian noise background," *IEEE Commun. Lett.*, vol. 23, no. 8, pp. 1369–1372, Aug. 2019.

[31] Q. Chen, X. Li, and Y. Wang, "SLA-driven cost-effective monitoring based on criticality for multi-tenant service-based systems," *IEEE Access*, vol. 6, pp. 48765–48775, 2018.

[32] S. Sengupta, S. Rayanchu, and S. Banerjee, "Network coding-aware routing in wireless networks," *IEEE-ACM Trans. Netw.*, vol. 18, no. 4, pp. 1158–1170, Aug. 2010.

[33] R. Chai, Z. Ma, C. Liu, and Q. Chen, "Service characteristics-oriented joint ACB, cell selection, and resource allocation scheme for heterogeneous M2M communication networks," *IEEE Syst. J.*, vol. 13, no. 3, pp. 2641–2652, Sep. 2019.

[34] S. Manjunath and G. Raina, "Stability and performance of compound TCP with a proportional integral queue policy," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 5, pp. 2139–2155, Sep. 2019.

[35] Q. Luo, X. Fang, Y. Sun, J. Ai, and C. Yang, "Self-learning hot data prediction: Where echo state network meets NAND flash memories," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 3, pp. 939–950, Mar. 2020.

[36] L. Guo, Y. Li, and Z. Cai, "Minimum-latency aggregation scheduling in wireless sensor network," *J. Combinat. Optim.*, vol. 31, no. 1, pp. 279–310, Jan. 2016.

[37] B. Malhotra, I. Nikolaidis, and M. A. Nascimento, "Aggregation convergecast scheduling in wireless sensor networks," *Wireless Netw.*, vol. 17, no. 2, pp. 319–335, Feb. 2011.

[38] M. Khan, S. Bashir, and A. Habib, "Semi round robin pairing and scheduling for uplink virtual multiple input multiple output (VMIMO) communications," *J. Spac. Technol.*, vol. 4, no. 1, pp. 61–66, Jul. 2014.

[39] D. Li, Q. Zhu, H. Du, and J. Li, "An improved distributed data aggregation scheduling in wireless sensor networks," *J. Combinat. Optim.*, vol. 27, no. 2, pp. 221–240, Feb. 2014.

[40] S. Boulkaboul, D. Djenouri, and N. Badache, "FDAP: Fast data aggregation protocol in wireless sensor networks," *Lect. Notes Comput. Sci.*, vol. 7469, no. 1, pp. 413–423, Jan. 2012.

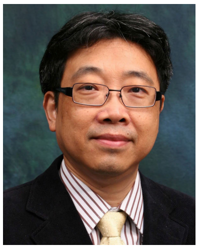
[41] J. Guo and Z. Zhang, "Research on location of chain convenience stores based on machine learning," in *Proc. IHMSC*, Hangzhou, China, Aug. 2019, pp. 225–228.

[42] X. Shao, C. Wang, C. Zhao, and J. Gao, "Traffic shaped network coding aware routing for wireless sensor networks," *IEEE Access*, vol. 6, pp. 71767–71782, 2018.

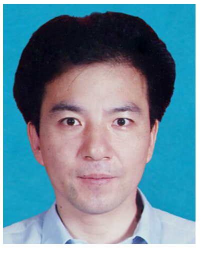
[43] X. Kong, F. Xia, Z. Ning, A. Rahim, Y. Cai, Z. Gao, and J. Ma, "Mobility dataset generation for vehicular social networks based on floating car data," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 3874–3886, May 2018.

[44] F. Lyu, N. Cheng, H. Zhou, W. Xu, W. Shi, J. Chen, and M. Li, "DBCC: Leveraging link perception for distributed beacon congestion control in VANETs," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4237–4249, Dec. 2018.

CHEN HUANG received the B.Eng. and Ph.D. degrees in communication and information system from the Huazhong University of Science and Technology, Wuhan, China, in 2005 and 2010, respectively. He is currently an Associate Professor with the Department of Computer and Information Engineering, Hubei University, Wuhan. His research interests include the Internet of Things, autonomous driving, machine learning, and big data analysis in brain–computer interface.



JIANNONG CAO (Fellow, IEEE) received the B.Sc. degree in computer science from Nanjing University, China, in 1982, and the M.Sc. and Ph.D. degrees in computer science from Washington State University, USA, in 1986 and 1990, respectively. He is currently a Chair Professor with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. He is also the Director of the Internet and Mobile Computing Laboratory, and the University Research Facility in Big Data Analytics. He has coauthored five books in *Mobile Computing* and *Wireless Sensor Networks*, co-edited nine books, and published over 600 articles in major international journals and conference proceedings. His research interests include parallel and distributed computing, wireless networks and mobile computing, big data and cloud computing, pervasive computing, and fault tolerant computing. He is a Distinguished Member of ACM and a Senior Member of China Computer Federation (CCF).



SHIHUI WANG received the B.Sc. degree in computer science from Wuhan University, Wuhan, China, in 1986, and the M.Sc. degree in software engineering from Zhengzhou University, Zhengzhou, China. He is currently a Professor with the Department of Computer and Information Engineering, Hubei University, Wuhan. He is also the Director of the Education Information Engineering Technology Research Center of Hubei Province. His research interests include big data analysis, artificial intelligence, and virtual reality. He is a Senior Member of China Computer Federation (CCF).

YAN ZHANG received the B.Sc. and M.Sc. degrees in computer science from Hubei University, Wuhan, China, in 1997 and 2002, respectively, and the Ph.D. degree in software engineering from Beihang University, Beijing, China. He is currently a Professor with the Department of Computer and Information Engineering, Hubei University. He is also the Vice Director of the Education Information Engineering Technology Research Center of Hubei Province. His research interests include information security, big data analysis, and software defect detection. He is a Senior Member of China Computer Federation (CCF).