

Article

# A Terminal-Oriented Distributed Traffic Flow Splitting Strategy for Multi-Service of V2X Networks

Fei Ding <sup>1,2,3,\*</sup> , Zitong Ma <sup>1</sup>, Zhiliang Li <sup>1</sup>, Ruoyu Su <sup>1</sup>, Dengyin Zhang <sup>1,2,3</sup> and Hongbo Zhu <sup>1,2,3</sup>

<sup>1</sup> Jiangsu Key Laboratory of Broadband Wireless Communication and Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; 15850567619@163.com (Z.M.); lizhiliang3@163.com (Z.L.); suruoyu@njupt.edu.cn (R.S.); zhangdy@njupt.edu.cn (D.Z.)

<sup>2</sup> Jiangsu Key Laboratory of Wireless Communications, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; zhb@njupt.edu.cn

<sup>3</sup> Key Laboratory of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

\* Correspondence: dingfei@njupt.edu.cn; Tel.: +86-25-83492013

Received: 16 April 2019; Accepted: 31 May 2019; Published: 6 June 2019



**Abstract:** With the development and the characteristics of terminal services of the 5G (5th-Generation network) Internet of Vehicles (IoVs), this paper proposes a distributed splitting strategy for multi-type services of 5G V2X (Vehicle to X) networks. Based on a service-oriented adaptive splitting strategy in heterogeneous networks, combined with various service types such as communications between the networks, terminal, and base stations, and the value-added services of 5G IoVs, the proposed strategy jointly considers delay and cost as optimization goals. By analyzing the characteristics of the different services, the proposed traffic flow splitting strategy is modeled as an optimization problem to efficiently split services in 5G V2X networks. The simulation results show that by setting the traffic distribution policy for each service, the distributed traffic flow splitting strategy can significantly improve network transmission efficiency and reduce the service costs in a vehicle V2X network.

**Keywords:** Internet of Vehicle; V2X network; traffic flow splitting strategy; distributed load sharing

## 1. Introduction

With the rapid development of wireless communication and network technologies, future wireless communication network systems will be heterogeneous networks that include coexistence, fusion, and complementation of multiple networks, such as cellular networks, WLANs (Wireless Local Area Networks), and WPANs (Wireless Personal Area Networks) [1,2]. In a 5G (5th-Generation network) vehicle networking scenario, each vehicle is a communication node, and the vehicle network will generate data of different dimensions, different structures, and different properties, which constitute the heterogeneity of the vehicle network [3]. Diversified vehicle networking services and a large number of terminals have put forward high throughput, ultra-low latency, and high stability requirements for the Internet of Vehicles [4,5].

The 5G communication technology applied in the Internet of Vehicles scenario has prompted vehicle networks to have more flexible architecture and new system elements (5G vehicle OBU (On Board Unit), 5G base stations, 5G mobile terminals, 5G cloud servers, etc.) [6]. In addition to V2X information interactions in the in-vehicle network, the vehicle to vehicle (V2V) network, and the vehicle mobile internet, a 5G vehicle network will also realize the interconnection between the OBU, base station, mobile terminal, and cloud server [7,8]. In such a hybrid network with multiple types of services, such as network heterogeneity, data heterogeneity, and different QoS (Quality of Service)

requirements, determining how to solve the network load, improve the utilization of wireless resources, and improve the quality of the vehicle networking service in the 5G scenario has become an important issue [9].

However, traditional wireless resource management only considers its own network conditions, maximizes the use of network resources, and provides “best effort” services, and therefore cannot adapt well to the development of the network. Multi-connection and parallel transmission will break through the limitations on the ability of a single network to provide the stable transmission of services, and give users high-quality transmission services [10,11]. These problems can be effectively coped with by utilizing traffic flow offloading technology, segmenting traffic flow data streams, and optimizing the communication traffic allocated to each network one by one [12]. Traffic flow diversion is a key part of 5G vehicle networking research. Through proper traffic flow distribution, traffic flow diversion can effectively alleviate the pressure of rapid growth of traffic volume [13].

5G vehicle networking terminals will carry V2X networking and communication services, as well as vehicle value-added services. Multi-type vehicle services and differentiated traffic service requirements pose a huge challenge to the traditional wireless resource management model. Therefore, in the 5G vehicle network scenario, the coordination of network resources, the optimization of the utilization of wireless resources, and the realization of resource allocation and management of multi-service functionality in a heterogeneous vehicle network is very important. These factors are especially important for the user experience of the vehicle information service. As a multi-attribute decision making approach, the analytical hierarchy process (AHP) efficiently decomposes complex problems into different levels and parts. The parts in the same level are compared and judged. Inspired by AHP, with the priority control based on the queueing theory, the main contributions of this paper are summarized as follows. We also introduce the fundamental ideas of AHP, the priority control, and queueing theory related to our work in Section 2.

- We establish the traffic flow splitting model of the multi-service distributed load sharing system to clearly present the 5G V2X communication scenario.
- We propose a traffic flow splitting algorithm to improve communication efficiency for 5G V2X communication scenario.

The remaining parts of this paper are organized as follows. In Section 2, we establish the fundamental architecture of V2X networks. We propose a traffic flow splitting algorithm in Section 3. We report the simulation results in Section 4 and conclude the paper in Section 5.

## 2. Related Works

At present, the research on the traffic flow shunting mechanism in a heterogeneous network environment mostly determines the optimization target according to the characteristics of the service data, resulting in the transformation of the traffic flow diversion problem into an optimization problem [14,15]. The above optimization target is then used to maximize or minimize processing to obtain an optimal service offload strategy. The traffic flow offloading mechanism is an effective way to solve the above problems by dividing the service data stream and optimizing the allocation to each wireless network one by one. This section will introduce the main processing ideas, and the advantages and disadvantages of current modes of distribution.

### 2.1. Analytic Hierarchy Process Based on Service Distribution Weight

The analytical hierarchy process (AHP) is a multi-attribute decision making method that combines qualitative analysis with quantitative analysis. Its basic idea is to decompose complex problems into several levels and elements, make simple comparisons and judgments between those elements at the same level, and finally, to derive the importance of the relationship between parameters [16].

In the IoT (Internet of Things)-based healthcare system, the multiple deployments of medical devices in a narrow area results in signal interference, and a hierarchical interface analysis (AHP)-based

network interface and channel selection algorithm have been proposed. The multi-channel MAC (Media Access Control) protocol in the networked ecosystem ultimately achieves low latency and high reliability goals [17]. A priority of different QoS requirements of heterogeneous IoT applications in a fog network has been proposed by enforcing the analytical framework using an analytic hierarchy process (AHP), then, a two-sided matching game to initiate stable association between the network infrastructure (i.e., fog devices) and IoT devices is formulated. Finally, the externalities in the matching game that occur due to job delay, along with the solution of the network resource allocation problem by applying a “best fit” resource allocation strategy during matching, are verified [18].

The AHP-based traffic flow offloading method needs to clarify the weight of each service, so that each service has a focus on the choice of network access. However, this scheme has its shortcomings. The AHP algorithm primarily allocates the weight of each service, but does not allocate the priority of each service. Therefore, it cannot meet the adaptive capability of service offloading in some scenarios, i.e., automatic network access and shunt control cannot be achieved.

### 2.2. Priority-Based Service Offloading

The priority-based service offloading scheme has priority control on services, and each service has a corresponding priority relationship and is stored in the service offload algorithm. Based on the current available resources, congestion level, and service characteristics of the network, the connection between the service flow and the network interface is automatically established, thereby replacing the user’s shunting decision, which can partially compensate for the deficiencies of the AHP algorithm. This allows services with higher priority to prioritize network resources in the case of heavy network load.

Service priority-based channel access technology (TP-CAT) using IEEE 802.15.6 protocol is proposed to minimize the transmission delay of critical data packets and to resolve collisions between other priority nodes during the back-off phase [19]. TP-CAT is a hybrid preemptive/non-preemptive recovery priority scheme that can handle the heterogeneity requirements of M2M (Machine-to-Machine) services while integrating into an LTE (Long Term Evolution) network [20]. In the IoT scenario, a class-based dynamic priority (CBDP) algorithm is proposed for delay sensing and scheduling management of uplink M2M services, and has minimal impact on the QoS of uplink H2H (Human-to-Human) services [21]. A constrained utility-optimization formulation of the joint-bandwidth allocation problem for multiple classes of traffic in inter-data center communication is proposed to handle priorities between traffic classes in a soft manner, and explicitly considers the delay requirement of interactive flows [22].

### 2.3. Queue-Based Service Shunting

Queuing theory can effectively analyze and simulate the dynamic behavior of data packets in a link. Multiple service data streams are queued in sequence to wait for processing by the processing unit, thereby improving data transmission efficiency. The processing unit does not consider the structure or characteristics of the traffic flow, i.e., it acts as a “black box.” Its function for data streams is to analyze processing and path selection, and finally to distribute the data stream to different links.

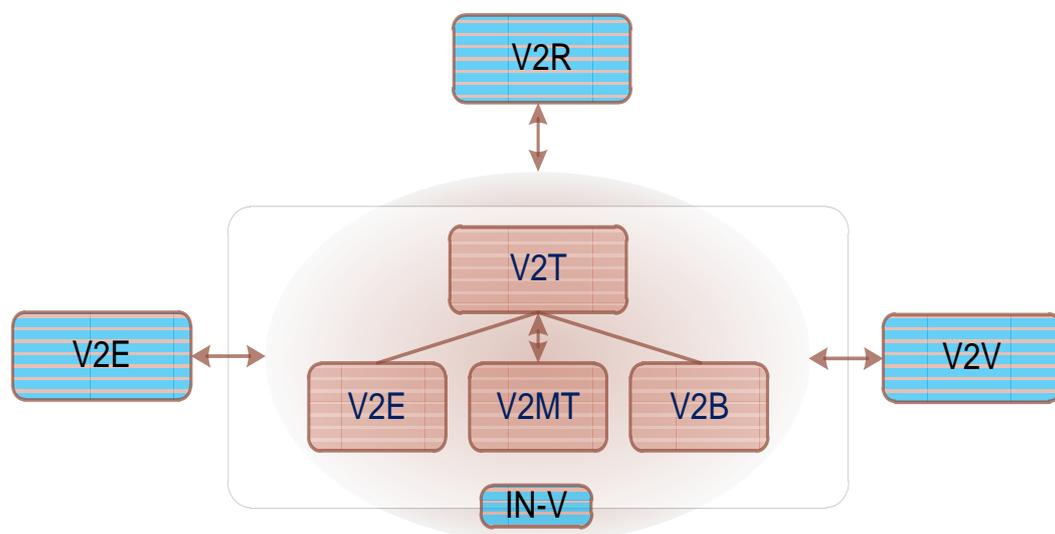
The adaptive traffic distribution mechanism based on a WLAN and wireless personal area network is studied in [23]. By modeling the transmission of data traffic between networks into a queuing system, the system transmission delay is minimized, thereby achieving load balancing [24,25]. Some literature [25] has also presented a probability-split algorithm that minimizes delay in heterogeneous wireless networks. Some literature [26] proposes an optimal distribution scheme based on dynamic traffic allocation in heterogeneous wireless network environments. This scheme aims to minimize the queue delay while maximizing the probability of a packet arriving at a given delay constraint. In order to support multiple users with multiple services, a resource allocation and burst scheduling algorithm for HPC (Highest Priority-Controlled)-DRX (Discontinuous Reception) /TRTS (Traffic Regulation plus Time Slicing) is introduced. The lower priority traffic never wakes up the user equipment (UE), and all incoming traffic of the UE is temporarily buffered at the evolved Node B (eNB), and then

transmitted in burst mode using the total bandwidth, resulting in better power-saving performance. The power-saving rate of the proposed mechanism is theoretically analyzed based on the M/G/1 busy period model [27].

Therefore, the analytic hierarchy process, priority control, and queuing methods for service shunting technology are always analyzed separately, which leads to a lack of a comprehensive solution. In this paper, we consider the advantages of the above methods and consider the service characteristics of the vehicular network terminal. Specifically, we first assign different weights according to different services. Then, the heterogeneous wireless sub-network is regarded as a plurality of different parallel queuing systems to model the service shunt problem as the joint optimization problem of delay and cost consumption. The proposed traffic flow splitting algorithm can achieve the optimization of transmission delay and service cost.

### 3. V2X Networking Architecture

The V2X communication scenario of vehicular networks is modeled in Figure 1, which is primarily divided into in-vehicle (IN-V) communication and out-of-vehicle communication (OUT-V). The IN-V communication is mainly based on the V2T (Vehicle to Terminal) installed in the car, which carries out vehicle environment (V2E) monitoring, and performs OBD (on-board diagnostics) diagnosis, in-vehicle video monitoring, positioning, and card swiping management, among other functions. It also carries out access communication and service of the mobile terminal in the vehicle (V2MT). V2B (Vehicle to Business) refers to the commercial information operation service in the vehicle. Extravehicular communication is also based on the vehicle terminal, which mainly implements environment awareness including V2P (Vehicle to Pedestrian) and V2I (Vehicle to Information) service requirements while driving, and realizes V2V communication with other vehicles to achieve safe driving.



**Figure 1.** Communication model in V2X Scenario.

In a 5G vehicle networking scenario, V2X communication includes V2E and V2MT communication in the vehicle, as well as V2E, V2V, and vehicle to road side) communication (V2R) outside the vehicle. Considering the demand for off-board communication of the vehicle V2X service, the traffic flow splitting strategy (TFSS) is run in each V2T terminal (software-based embedded application), and the TFSS is managed by the access controller (AC) cloud, as illustrated in Figure 2. In order to better support the vehicle information service, the AC cloud needs to connect to the portal server and the user authentication (UA) server, provide content to the in-vehicle information service, and verify the service user. The business operation support system (BOSS) conducts operational billing and customer management for vehicle services.

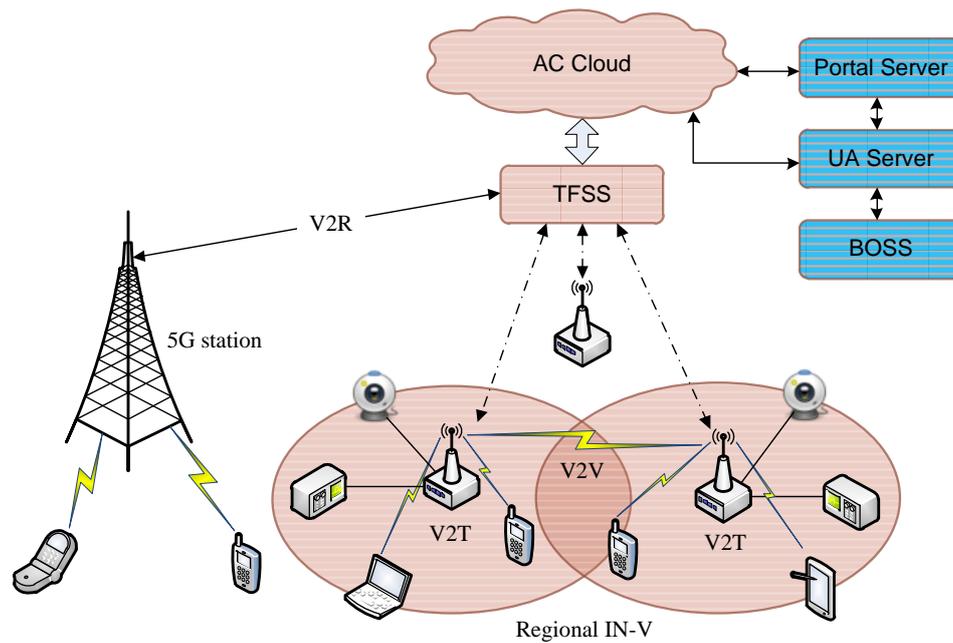


Figure 2. V2X-based service architecture.

#### 4. TFSS Algorithm

A typical V2X communication system consists of three parts: the transmitting end, the heterogeneous network, and the receiving end. The heterogeneous network integrates various types of networks, such as WLAN and cellular mobile communication networks. The development of in-vehicle communication will cause congestion in IEEE 802.11p standard communication. Therefore, it is necessary to improve communication efficiency by the traffic flow splitting of the vehicle terminal (V2T). The specific characterization of this is that the user selects different transmission networks that correspond to different bandwidths and costs.

##### 4.1. System Model

A traffic flow splitting model for multiple types of vehicle services is provided in Figure 3. The offloading model is carried by the in-vehicle gateway and is responsible for transmitting or receiving the communication data stream. According to the V2X communication scenario of the Internet of Vehicles, the vehicle communication network includes various types of service requirements, such as V2V, V2E, V2MT, and V2R. For each type of in-vehicle service, the data stream is sequentially arranged into  $n$  queues on the transmitting end to wait for the offload processing. Nodes A, B, and C are equivalent to the data stream interfaces of each type of service accessing the uplink network. Therefore, if the in-vehicle gateway knows the available bandwidth of each uplink access network and the characteristics of the in-vehicle service that needs to be transmitted, it is easier to control the traffic ratio flowing into each communication link, thereby resulting in traffic offload in the in-vehicle heterogeneous network scenario. The data flow for each type of in-vehicle service conforms to the Poisson distribution (the parameter of the Poisson distribution is  $\lambda$ ).

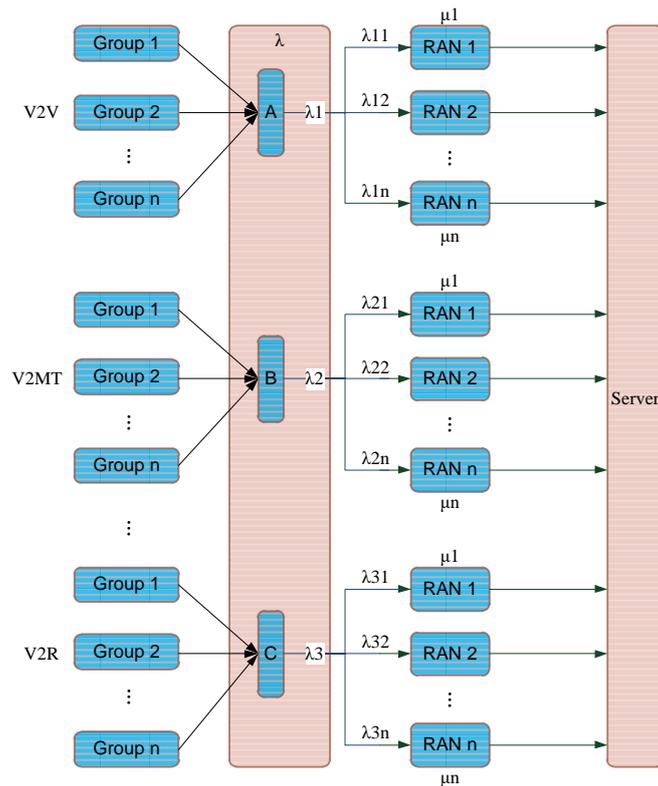


Figure 3. Multi-service distributed load sharing model.

Suppose the total service data arrival rate in the network is  $R$  (bps), the arriving data packet obeys the Poisson distribution with the parameter  $\lambda$ , and the length of the data packet is  $L_p$ , then  $\lambda = R/L_p$ . The data rate relationship between the data traffic in the total in-vehicle network and the data traffic on each subnet can be expressed as:  $R = \sum_{i=1}^n R_i$ . Because  $\lambda_i = R_i/L_p$ , the equation can also be written as  $\lambda = \sum_{i=1}^n \lambda_i$ . Suppose  $\mu_i$  is the  $i^{\text{th}}$  wireless communication technology (RAT), that is, the service rate of the  $i^{\text{th}}$  radio link, and that the number of available network resources of the  $i^{\text{th}}$  RAT is  $R_a^i$  (bps), then  $\mu_i = R_a^i/L_p$ . Because this strategy uses delay and cost as joint optimization goals, we need to find the expression of the delay.

According to queuing theory, the state probability  $P_n(t)$  of  $n$  packets in the system satisfies the following equilibrium equation:

$$\begin{cases} \lambda P_0 = \mu P_1 \\ \lambda P_{n-1} + \mu P_{n+1} = (\lambda + \mu) P_n \end{cases} \tag{1}$$

The state probability  $P_n(t)$  is satisfied as follows:

$$\begin{cases} P_1 = \frac{\lambda}{\mu} P_0 \\ P_2 = \frac{\lambda}{\mu} P_1 = \left(\frac{\lambda}{\mu}\right)^2 P_0 \\ \dots \\ P_n = \frac{\lambda}{\mu} P_{n-1} = \left(\frac{\lambda}{\mu}\right)^n P_0 \end{cases} \tag{2}$$

It can be seen from Equation (2) that the state probability  $P$  is a traversed Markov chain, i.e., the following state is satisfied: the next state is only related to the current state, and the state transition of the next state is independent of the value of the current state, that is, the transition is fixed.

**Theorem 1.** In the queuing theory model, if the arrival rate  $\lambda$  of the packet and the service rate  $\mu$  satisfy  $\lambda < \mu$ , the output queue also obeys the Poisson distribution with the parameter  $\lambda$ , and the queue maintains a stable queuing state.

Theorem 1 can also be expressed as: corresponding to traffic flow within the communication network, if  $\lambda > \mu$ , that is, if the traffic is larger than the available bandwidth, serious congestion will occur. Therefore, it is necessary to ensure  $\lambda < \mu$ .

Because there are hidden conditions, the sum of the system state probabilities is 1, i.e.,

$$\sum_{n=1}^{\infty} P_n = P_0 \left[ 1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2 + \dots + \left(\frac{\lambda}{\mu}\right)^n \right] = P_0 \cdot \frac{1}{1 - \frac{\lambda}{\mu}} = 1 \tag{3}$$

Because

$$P_0 = 1 - \frac{\lambda}{\mu} \tag{4}$$

then the state probability  $P_n$  can be expressed as

$$P_n = \left(\frac{\lambda}{\mu}\right)^n \cdot \left(1 - \frac{\lambda}{\mu}\right) \tag{5}$$

Suppose  $L_{s_i}$  is the average number of packets in the  $i^{\text{th}}$  network and satisfies as

$$L_{s_i} = \sum_{n=0}^{\infty} n \cdot P_n = \sum_{n=0}^{\infty} n \cdot \left(\frac{\lambda_i}{\mu_i}\right)^n \left(1 - \frac{\lambda_i}{\mu_i}\right) = \frac{\lambda_i}{\mu_i - \lambda_i} \tag{6}$$

and the average service waiting time, that is, the transmission delay  $D_i$ , can be expressed as

$$D_i = \frac{L_{s_i}}{\lambda_i} = \frac{1}{\mu_i - \lambda_i} \tag{7}$$

then the total transmission delay,  $D(\lambda)$ , in the system is

$$D(\lambda) = \sum_{i=1}^n D_i = \sum_{i=1}^n \frac{1}{\mu_i - \lambda_i} \tag{8}$$

$\theta_i$  denotes the transmission cost required for the unit number of packets to be transmitted on the  $i^{\text{th}}$  network. In fact, if the packet length and the cost per bit of data transmission are known, then  $\theta_i$  can be found by their multiplication. Both the packet length and cost per bit of data are known in actual network transmission, i.e.,  $\theta_i$  is a known amount for the entire network [6].

The cost consumption on the  $i^{\text{th}}$  RAT can be expressed as

$$C_i = L_{s_i} \cdot \theta_i = \frac{\lambda_i \cdot \theta_i}{\mu_i - \lambda_i} \tag{9}$$

Therefore, the total cost consumption  $C(\lambda)$  in a heterogeneous network is

$$C(\lambda) = \sum_{i=1}^n C_i = \sum_{i=1}^n \frac{\lambda_i \cdot \theta_i}{\mu_i - \lambda_i} \tag{10}$$

Therefore, the joint optimization goal of the service offload problem of the in-vehicle gateway, that is, the utility function, can be expressed as

$$F(\lambda) = \eta_1 D(\lambda) + \eta_2 C(\lambda) = \eta_1 \sum_{i=1}^n \frac{1}{\mu_i - \lambda_i} + \eta_2 \sum_{i=1}^n \frac{\lambda_i \cdot \theta_i}{\mu_i - \lambda_i} \tag{11}$$

where  $\eta_1$  and  $\eta_2$  are weighting factors, and  $\eta_1 + \eta_2 = 1$ . If the gateway service has a high requirement for the delay, that is, the time-sensitive service, then  $\eta_1 > \eta_2$ . If the gateway service can tolerate a higher delay and has a higher cost requirement, that is, a cost-type service, then  $\eta_2 > \eta_1$ .

$D(\lambda)$  represents the total time consumption and  $C(\lambda)$  represents the total cost. Therefore, the less the delay and cost, the higher the gain. In other words, the joint optimization of the service offload of the in-vehicle gateway translates into a solution for the minimization problem.

#### 4.2. Optimization Problem

According to the optimization goal of the service offload, the utility function is established as follows:

$$F(\lambda) = \eta_1 D(\lambda) + \eta_2 C(\lambda) = \eta_1 \sum_{i=1}^n \frac{1}{\mu_i - \lambda_i} + \eta_2 \sum_{i=1}^n \frac{\lambda_i \cdot \theta_i}{\mu_i - \lambda_i} \tag{12}$$

In the above equation,  $\eta_1$  and  $\eta_2$  are weighting factors, and  $\eta_1 + \eta_2 = 1$ . If the gateway service requires a short delay, that is, a delay-sensitive service, then  $\eta_1 > \eta_2$ . If the gateway service can tolerate a higher delay and requires a low cost, that is, a cost-based service, then  $\eta_2 > \eta_1$ . When  $\eta_1 = 1, \eta_2 = 0$ , only the system delay is considered. Conversely, when  $\eta_1 = 0, \eta_2 = 1$ , only the cost is considered at the time. Therefore, the specific values of  $\eta_1$  and  $\eta_2$  depend on the characteristics of the gateway service.

The traffic flow splitting algorithm will be affected by the stability of the network transmission, the characteristics of the service itself, and other factors in the network. Therefore, the following should be set: (1) the transmission to greater system stability requirements are  $0 \leq \lambda_i < \mu_i$ , that is, the service rate of the data packet is better guaranteed than the arrival rate and (2) the delay limit is  $D_i = \frac{1}{\mu_i - \lambda_i} < D_0$ , where  $D_0$  is the maximum delay of the system.

Thus, the problem is equivalent to the following optimization problem with practical constraints:

$$\begin{cases} \min(F(\lambda)) \\ \lambda = \sum_{i=1}^n \lambda_i \\ 0 \leq \lambda_i < \mu_i (i = 1, 2, \dots, n) \\ \frac{1}{\mu_i - \lambda_i} < D_0 \end{cases} \tag{13}$$

The Lagrange multiplier method is combined with KKT (Karush-Kuhn-Tucher) to solve the optimization problem, and the global optimal solution is obtained as follows:

The Lagrange multipliers,  $\omega_0$  and  $\omega_i$ , are introduced to construct a Lagrangian function:

$$L(\lambda, \omega) = \eta_1 \sum_{i=1}^n \frac{1}{\mu_i - \lambda_i} + \eta_2 \sum_{i=1}^n \frac{\lambda_i \cdot \theta_i}{\mu_i - \lambda_i} - \omega_0 \left( \sum_{i=1}^n \lambda_i - \lambda \right) - \sum_{i=1}^n \omega_i \left( \mu_i - \max \left( \lambda_i, \mu_i - \frac{1}{T_{\max} - T_j} \right) \right) \tag{14}$$

defined as  $\varphi_1(\lambda) = \sum_{i=1}^n \lambda_i - \lambda$ ,  $\varphi_i(\lambda) = \sum_{i=1}^n \omega_i \left( \mu_i - \max \left( \lambda_i, \mu_i - \frac{1}{T_{\max} - T_j} \right) \right)$ .

According to KKT conditions,

$$\omega_i \cdot \varphi_i = 0 \quad (i = 0, 1, \dots, n) \tag{15}$$

Therefore, Equation (14) can be simplified to

$$L(\lambda, \omega) = \eta_1 \sum_{i=1}^n \frac{1}{\mu_i - \lambda_i} + \eta_2 \sum_{i=1}^n \frac{\lambda_i \cdot \theta_i}{\mu_i - \lambda_i} - \omega_0 \left( \sum_{i=1}^n \lambda_i - \lambda \right) \tag{16}$$

Because the KKT condition satisfies the function  $L$ , the derivation for each  $\lambda$  is 0,

$$\begin{cases} \frac{\partial L}{\partial \lambda_i} = 0 \\ \frac{\partial L}{\partial \omega_0} = 0 \end{cases} \tag{17}$$

Then,

$$\begin{cases} \frac{\eta_1}{(\mu_1 - \lambda_1)^2} + \frac{\eta_2 \mu_1 \theta_1}{(\mu_1 - \lambda_1)^2} - \omega_0 = 0 \\ \frac{\eta_1}{(\mu_2 - \lambda_2)^2} + \frac{\eta_2 \mu_2 \theta_2}{(\mu_2 - \lambda_2)^2} - \omega_0 = 0 \\ \dots \\ \frac{\eta_1}{(\mu_n - \lambda_n)^2} + \frac{\eta_2 \mu_n \theta_n}{(\mu_n - \lambda_n)^2} - \omega_0 = 0 \\ \lambda_1 + \lambda_2 + \dots + \lambda_n = \lambda \end{cases} \tag{18}$$

Assuming that there are two heterogeneous subnets in the network, i.e.,  $n = 2$ , then Equation (18) can be reduced to the following:

$$\begin{cases} \frac{\eta_1}{(\mu_1 - \lambda_1)^2} + \frac{\eta_2 \mu_1 \theta_1}{(\mu_1 - \lambda_1)^2} - \omega_0 = 0 \\ \frac{\eta_1}{(\mu_2 - \lambda_2)^2} + \frac{\eta_2 \mu_2 \theta_2}{(\mu_2 - \lambda_2)^2} - \omega_0 = 0 \\ \lambda_1 + \lambda_2 = \lambda \end{cases} \tag{19}$$

Next,

$$\begin{cases} \lambda_1 = \mu_1 - (\mu_1 + \mu_2 - \lambda) \cdot \frac{\sqrt{\eta_1 + \eta_2 \mu_1 \theta_1}}{\sqrt{\eta_1 + \eta_2 \mu_1 \theta_1} + \sqrt{\eta_1 + \eta_2 \mu_2 \theta_2}} \\ \lambda_2 = \mu_2 - (\mu_1 + \mu_2 - \lambda) \cdot \frac{\sqrt{\eta_1 + \eta_2 \mu_2 \theta_2}}{\sqrt{\eta_1 + \eta_2 \mu_1 \theta_1} + \sqrt{\eta_1 + \eta_2 \mu_2 \theta_2}} \end{cases} \tag{20}$$

When there are  $n$  heterogeneous subnets in the system, the arrival rate of the data packet is

$$\lambda_i = \mu_i - (\mu_1 + \mu_2 + \dots + \mu_n - \lambda) \cdot \frac{\sqrt{\eta_1 + \eta_2 \mu_i \theta_i}}{\sqrt{\eta_1 + \eta_2 \mu_1 \theta_1} + \sqrt{\eta_1 + \eta_2 \mu_2 \theta_2} \dots + \sqrt{\eta_1 + \eta_2 \mu_n \theta_n}} \tag{21}$$

For the multiple types of services in V2X communication, the two-level service offloading scheme (as shown in Figure 3) can further realize traffic flow splitting. The optimization goal is the joint optimization of delay and energy consumption is as follows:

$$F(\lambda) = \eta_1 D(\lambda) + \eta_2 C(\lambda) = \eta_1 \sum_{i=1}^n \frac{1}{\mu_i - \lambda_i} + \eta_2 \sum_{i=1}^n \frac{\lambda_i \cdot \theta_i}{\mu_i - \lambda_i} \tag{22}$$

## 5. Simulation Results

### 5.1. V2X Service Classification

V2X services are divided into many types and collected from multiple sources, as presented in Table 1. According to the further classification of various services in the V2X communication scenario, and combining the vehicle communication modes corresponding to the services, it can be seen that each service also includes multiple vehicle communication modes. For example, V2V communication between vehicles includes both V2V communication and V2R communication between the vehicle and the base station. In other words, related information such as communication and location between vehicles needs to be reported in real time and shared within a certain area. The V2X service is mainly carried by the V2T. Therefore, the traffic flow control and management of the V2X service can further optimize the communication efficiency of the V2T terminal.

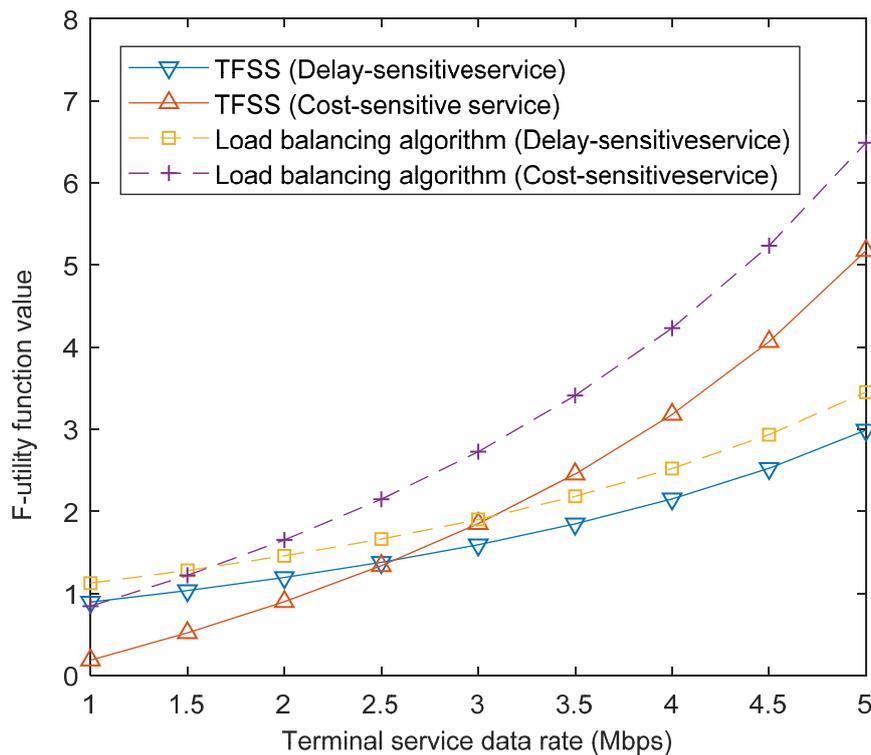
**Table 1.** V2X-based service classification.

Service Scenario	Service Classification	V2X Service Mode
IN-V	Driving behavior data	V2B + V2R
	Paying behavior data	V2B + V2R
	Travel behavior data	V2B + V2R
	Vehicle info data	V2E (IN-V) + V2R
	In-vehicle monitoring data	V2E (IN-V) + V2R
	Vehicle real-time location data	V2B + V2R
	Vehicle operation data	V2B + V2R
OUT-V	Vehicle to pedestrian	V2E (OUT-V)
	Vehicle to traffic light	V2E (OUT-V)
	Crowdsourcing road condition data	V2E (OUT-V)
	Out-of-vehicle monitoring data	V2E (OUT-V)
V2V	Vehicle to vehicle communication	V2V + V2R
User service	In-vehicle information service (information, games, internet video, etc.)	V2R + V2MT + V2B
	Pushing cloud information (V2T and user terminals connected to V2T)	V2R + V2MT + V2B
Road data service	High resolution satellite images	V2E + V2R
	Aerial photography data	V2E + V2R
	Road infrastructure data	V2E + V2R

### 5.2. Service Shunting Scheme Based on Queuing Theory

In the heterogeneous network scenario, the performance of the adaptive shunting algorithm is simulated and analyzed for different characteristics of the vehicle network service, then compared with the load balancing algorithm [24] to verify the feasibility and effectiveness of the proposed TFSS algorithm.

Figure 4 compares the service offload performance results when the load balancing and traffic flow splitting algorithm are applied to delay-sensitive services and cost-sensitive services, respectively. The delay-sensitive service assumes  $\eta_1 = 0.8$ ,  $\eta_2 = 0.2$ . The cost-sensitive service assumes  $\eta_1 = 0.2$ ,  $\eta_2 = 0.8$ . It is also assumed that there are three data links in the heterogeneous network, that is,  $n = 3$ , and their available resources are 2 Mb/s, 3 Mb/s, and 4 Mb/s, respectively. The network cost is 3, 2, and 1, respectively.



**Figure 4.** Utility value comparison results.

In Figure 4, the smaller the value of the utility function, the better the shunt energy efficiency. The proposed adaptive shunting algorithm is feasible for both delay-sensitive and cost-sensitive services, and the service offload performance is better than the load balancing algorithm. Moreover, with the increase of the speed requirement of the V2X services, the value of the F-utility function value clearly increases. This indicates that as the data rate of the terminal service increases, the service demand is allocated to three wireless links, that is, the algorithm is effective and the terminal service is shunted. It can also be seen from Figure 4 that when the service rate demand of the terminal is low, the cost-sensitive service will achieve better system energy efficiency, i.e., will have a lower F-utility function value. When the value of the V2X service data is increased to about 2.5 Mbps, the system of the delay-sensitive service is more efficient than the cost-sensitive service. Therefore, we believe that when there are more service requirements in V2X heterogeneous networks, delay is a key factor in determining system energy efficiency compared to cost factors.

The system delay comparison of the proposed adaptive service shunting algorithm in different scenarios of V2X service rate requirements is provided in Figure 5. The system delay of the adaptive shunting strategy is always lower than the load balancing algorithm. This is because the adaptive shunting algorithm accounts for the delay factor in the system, and the load balancing algorithm does not consider the service characteristics (assuming the system has the same transmission delay). It is also apparent that the system delay of the delay-sensitive service is always lower than the cost-sensitive service, which also reflects that the adaptive shunting algorithm can effectively implement service offload when it is oriented to specific V2X service requirements. Moreover, when the terminal service rate requirement is greater than 4 Mbps, the system delay starts to increase rapidly. This is primarily because when traffic shunting occurs in a heterogeneous network, system congestion will also occur, which affects the service delay.

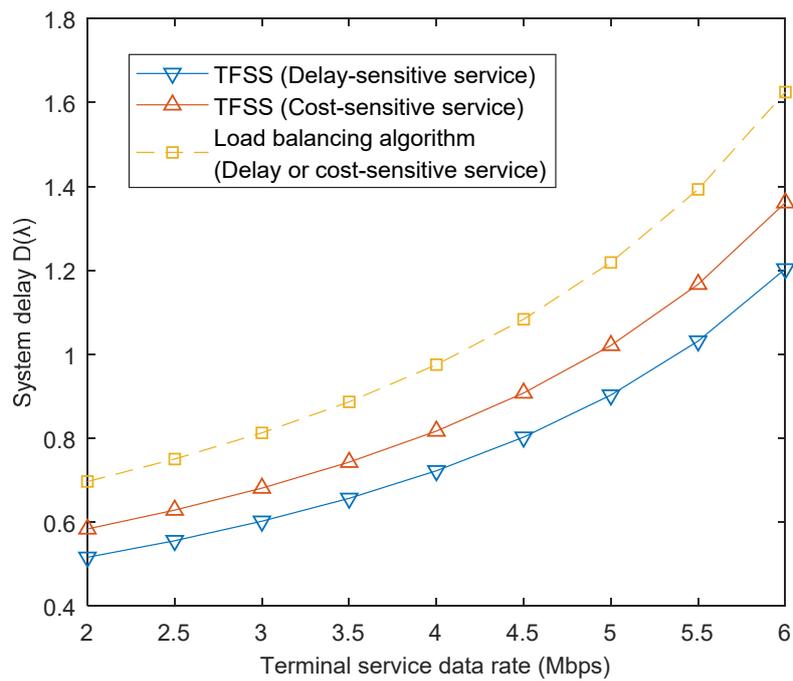


Figure 5. System delay comparison results.

Figure 6 shows the comparison of system cost consumption of terminal services in different rate demand scenarios. The system cost consumption of the adaptive shunting algorithm is significantly lower than that of the load balancing algorithm, and as the terminal traffic increases, the trend effect becomes increasingly obvious. Moreover, the cost of the adaptive shunting algorithm for cost-sensitive services is lower than that of delay-sensitive services. This indicates that the adaptive shunting algorithm has certain advantages in dealing with network tariff consumption in the distribution of traffic through different links.

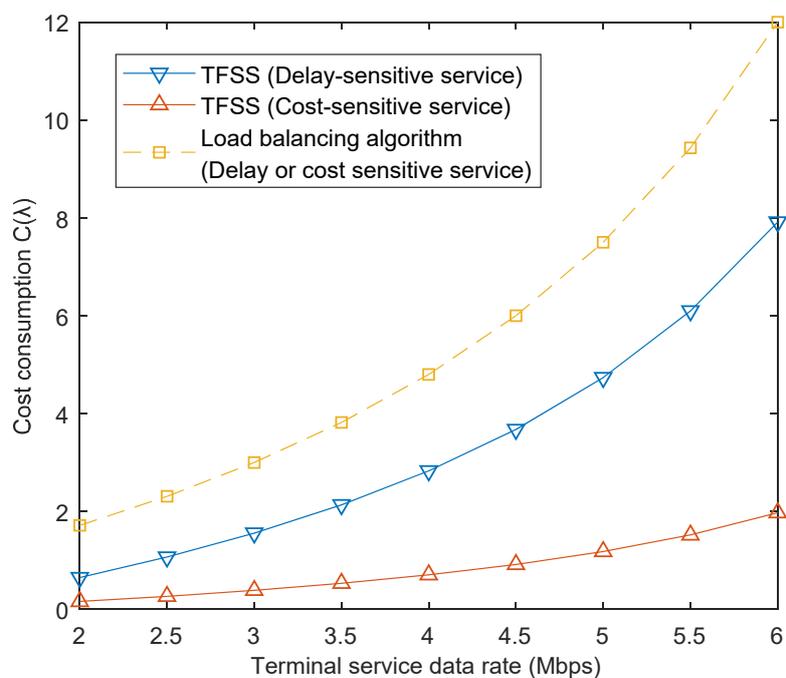


Figure 6. Total system cost comparison results.

### 5.3. Two-Level Distributed Offloading Scheme

In this section, we develop a performance analysis of the traffic flow distribution scheme for multi-service of the V2X network by formulating a specific 5G networking application scenario, and compare the proposed scheme with the adaptive shunting strategy for a single type of service.

V2Ts in operating vehicles (such as public buses) in cities improve the ability to operate content by expanding local storage. The main service types of V2Ts include high-bandwidth services such as online music, games, and video. The interactive service requirements of the vehicle terminal and the 5G base station are relatively small. The simulation parameter settings in this scenario are presented in Table 2. This scenario adapts to urban traffic sections, and guarantees vehicular information services (in-vehicle of V2T, V2MT, and V2B services), as well as out-of-vehicle road condition perception (V2E and V2R services out-of-vehicle), and belongs to the service type with high delay requirements.

Figure 7 shows the comparison results of delay-sensitive service utility values for multi-type service offloading and single-type service offloading when the service rate requirements of the terminal are different. When the terminal service rate demand is low, the F-utility value of the single type service split is better. When the terminal service rate demand exceeds about 3.8 Mbps, the F-utility value of the multi-type, multi-service split algorithm is higher. The performance optimization effect is more obvious, i.e., the classification performance of the adaptive shunt algorithm is better reflected.

Table 2. Multi-service splitting parameters (1).

Service Type	Service Characteristics	Service Proportion
V2T (IN-V)	Delay-sensitive (real-time)	10%
V2MT (IN-V)		20%
V2B (IN-V)		20%
V2E (OUT-V)		10%
V2R(OUT-V)		20%
V2E (IN-V)	Cost-sensitive (not real-time)	10%
V2V (OUT-V)		10%

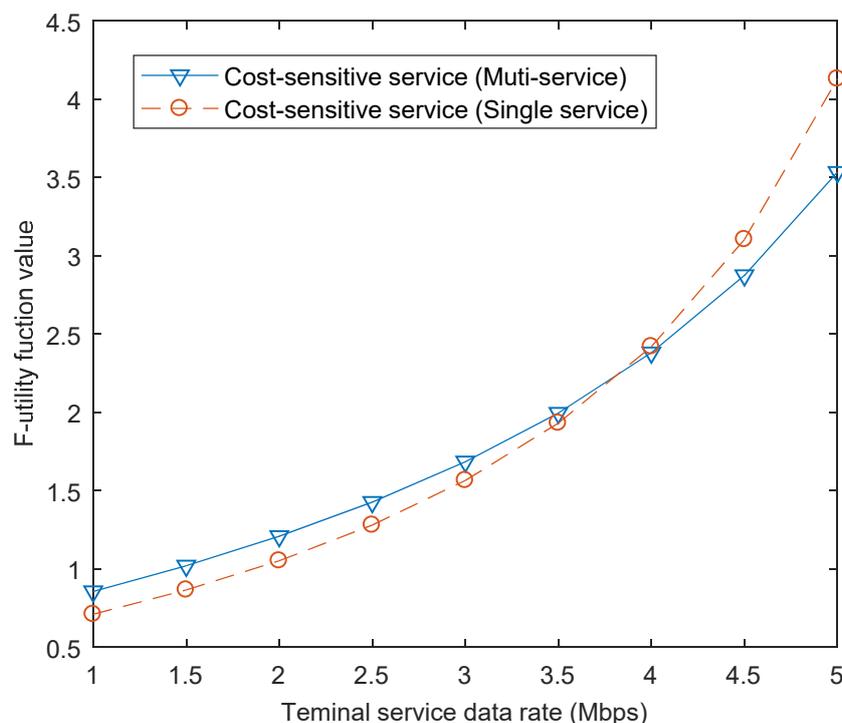


Figure 7. Delay-sensitive service utility value comparison results.

When the vehicle terminal in the Internet of Vehicles needs to communicate directly with the 5G base station to obtain its real-time location, traffic conditions, online video, etc., the service data generated by the vehicle terminal is mainly cost-sensitive. The simulation parameters are set as listed in Table 3, and the simulation results are shown in Figure 8. This scenario is suitable for expressway areas. In this scenario, priority should be given to vehicle condition monitoring, perception, and communication services involved in safe driving (which belong to delay-sensitive services). Large-bandwidth data services such as user information services (in-vehicle V2MT and V2B services) and in-vehicle security monitoring (in-vehicle V2E service) are cost-sensitive services.

Table 3. Multi-service splitting parameters (2).

Service Type	Service Characteristics	Service Proportion
V2T (IN-V)	Delay-sensitive (real-time)	5%
V2E (OUT-V)		5%
V2R (OUT-V)		5%
V2V (OUT-V)		5%
V2MT (IN-V)	Cost-sensitive (not real-time)	30%
V2B (IN-V)		20%
V2E (IN-V)		30%

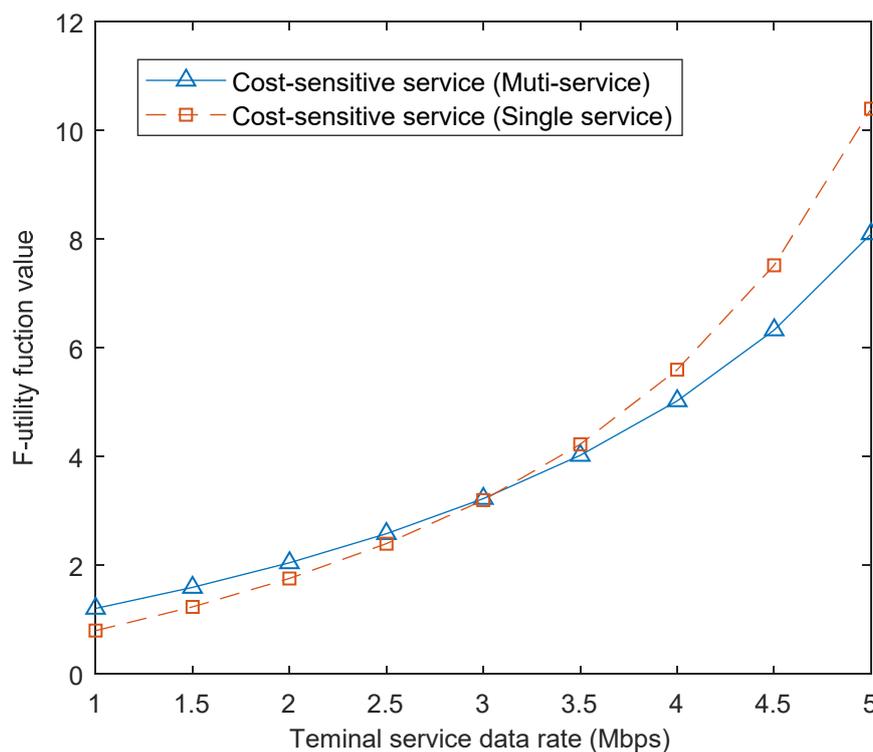


Figure 8. Cost-sensitive service utility value comparison results.

The load-separation results of the cost-sensitive service in both the single-type and multi-type service scenarios when the service data rate requirements of the vehicle-mounted terminal are different are shown in Figure 8. When the service rate requirement of the vehicle terminal is lower than 3 Mbps, the traffic splitting effect for a single type of service is relatively good. However, when the service rate demand exceeds 3 Mbps, the F-utility value of the multi-type service shunt is better, and the utility effect becomes increasingly obvious.

## 6. Conclusions

This paper examined the service requirements and communication modes of 5G vehicle networking terminals. As the typical service mode of 5G vehicular networks, a V2X-based vehicle service model is proposed. On this basis, the distributed traffic flow splitting strategy adapted to multi-service of V2X networks is studied. In the service classification process, multiple nodes are selected as the interfaces for the corresponding service data streams flowing into the sub-network in the V2X scenario. Then, each node distributes the data streams to the heterogeneous sub-networks to implement the offload transmission according to the adaptive offloading algorithm. The traffic flow splitting strategy takes delay and cost as joint optimization goals, and realizes the combination of the single service offloading scheme and the prioritized multi-service diversion scheme. Throughout the research in this paper, we realized that the following aspects need to be studied: (1) based on the V2X networking scenario, the multi-service shunting algorithm combined with vehicle mobility needs further study and (2) this paper mainly studies the parallel multi-network transmission for single-vehicle terminals, and does not consider the situation of regional multi-terminals. These are the directions for further research.

**Author Contributions:** F.D. completed the methodology; F.D., Z.M. and Z.L. completed the design; Z.M. and Z.L. wrote the paper. R.S. revised the paper. D.Z. and H.Z. reviewed the paper.

**Funding:** This work is partially supported by the National Natural Science Foundation of China (Nos. 61427801 and 61872423), the Ministry of Education-China Mobile Research Foundation, China (No. MCM20170205), the Scientific Research Foundation of the Higher Education Institutions of Jiangsu Province, China (No. 17KJB510043), the Six Talent Peaks project in Jiangsu Province (No. DZXX-008), and the Research Foundation for Advanced Talents of Nanjing University of Posts and Telecommunications (No. NY217146).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Guiomar, F.P.; Alimi, I.A.; Monteiro, P.P.; Gameiro, A. Flexible Infrastructure for the Development and Integration of Access/Fronthauling Solutions in Future Wireless Systems. In Proceedings of the IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Kalamata, Greece, 25–28 June 2018; pp. 1020–1024.
2. Wang, C.J.; Hsu, C.W. A microstrip wideband monopole antenna for multisystem integration by utilizing stepped-impedance structure and L-shaped slot. *Int. J. RF Microw. Comput. Aided Eng.* **2017**, *27*, 1–12. [[CrossRef](#)]
3. Sherazi, H.H.R.; Khan, Z.A.; Iqbal, R. A Heterogeneous IoV Architecture for Data Forwarding in Vehicle to Infrastructure. *Commun. Mob. Inf. Syst.* **2019**, *2019*, 3101276a. [[CrossRef](#)]
4. Sutagundar, A.V.; Attar, A.H.; Hatti, D.I. Resource Allocation for Fog Enhanced Vehicular Services. *Wirel. Pers. Commun.* **2019**, *104*, 1473–1491. [[CrossRef](#)]
5. Labiod, M.A.; Gharbi, M.; Coudoux, F.X.; Corlay, P.; Doghmane, N. Enhanced adaptive cross-layer scheme for low latency HEVC streaming over Vehicular Ad-hoc Networks (VANETs). *Veh. Commun.* **2019**, *15*, 28–39. [[CrossRef](#)]
6. Wang, L.M.; Liu, X.L. NOTSA: Novel OBU with Three-Level Security Architecture for Internet of Vehicles. *IEEE Internet Things J.* **2018**, *5*, 3548–3558. [[CrossRef](#)]
7. Saleem, M.A.; Zhou, S.J.; Sharif, A. Data Transmission Using IoT in Vehicular Ad-Hoc Networks in Smart City Congestion. *Mob. Netw. Appl.* **2019**, *24*, 248–258. [[CrossRef](#)]
8. Ahmed, A.A.; Alzahrani, A.A. A comprehensive survey on handover management for vehicular ad hoc network based on 5G mobile networks technology. *Trans. Emerg. Telecommun. Technol.* **2019**, *30*, 1–19. [[CrossRef](#)]
9. Kuutti, S.; Fallah, S.; Katsaros, K.; Dianati, M.; McCullough, F.; Mouzakitis, A. A Survey of the State-of-the-Art Localization Techniques and Their Potentials for Autonomous Vehicle Applications. *IEEE Internet Things J.* **2018**, *5*, 829–846. [[CrossRef](#)]
10. Li, J.; Zhang, X.; Wang, S.; Wang, W. Context-Oriented Multi-RAT User Association and Resource Allocation with Triple Decision in 5G Heterogeneous Networks. *China Commun.* **2018**, *15*, 72–85. [[CrossRef](#)]

11. Pokhrel, S.R.; Panda, M.; Vu, H.L. Analytical Modeling of Multipath TCP Over Last-Mile Wireless. *IEEE ACM Trans. Netw.* **2017**, *25*, 1876–1891. [[CrossRef](#)]
12. Kim, B.H.; Calin, D. On the Split-TCP Performance over Real 4G LTE and 3G Wireless Networks. *IEEE Commun. Mag.* **2017**, *55*, 124–131. [[CrossRef](#)]
13. Li, G.Z.; Guo, S.T.; Yang, Y. Traffic Load Minimization in Software Defined Wireless Sensor Networks. *IEEE Internet Things J.* **2018**, *5*, 1370–1378. [[CrossRef](#)]
14. Zhao, H.; Zheng, Q.H.; Zhang, W.Z.; Du, B.; Li, H. A Segment-Based Storage and Transcoding Trade-off Strategy for Multi-version VoD Systems in the Cloud. *IEEE Trans. Multimed.* **2017**, *19*, 149–159. [[CrossRef](#)]
15. Oguejiofor, O.S.; Zhang, L.X.; Nawaz, N. UE-centric clustering and resource allocation for practical two-tier heterogeneous cellular networks. *IET Commun.* **2018**, *12*, 2384–2392. [[CrossRef](#)]
16. Drissi, M.; Oumsis, M.; Aboutajdine, D. A Multi-Criteria Decision Framework for network selection over LTE and WLAN. *Eng. Appl. Artif. Intell.* **2017**, *66*, 113–127. [[CrossRef](#)]
17. Kim, B.S.; Kim, S. An AHP-Based Interface and Channel Selection for Multi-channel MAC Protocol in IoT Ecosystem. *Wirel. Pers. Commun.* **2017**, *93*, 97–118. [[CrossRef](#)]
18. Abedin, S.F.; Alam, M.G.R.; Kazmi, S.M.A.; Tran, N.H.; Niyato, D.; Hong, C.S. Resource Allocation for Ultra-Reliable and Enhanced Mobile Broadband IoT Applications in Fog Network. *IEEE Trans. Commun.* **2019**, *67*, 489–502. [[CrossRef](#)]
19. Ambigavathi, M.; Sridharan, D. Traffic Priority Based Channel Assignment Technique for Critical Data Transmission in Wireless Body Area Network. *J. Med Syst.* **2018**, *42*, 206. [[CrossRef](#)]
20. AlQahtani, S.A. Analysis of a Hybrid Priority Scheme for Multiclass M2M Communications over LTE-A Networks. *Arab. J. Sci. Eng.* **2018**, *43*, 6981–6993. [[CrossRef](#)]
21. Giluka, M.K.; Pasca, T.V.; Rajoria, N.; Tamma, B.R. On Efficient Scheduling of H2H Traffic and Reducing Signaling Overhead due to Uplink Small Data M2M Traffic in LTE-A Networks. *Wirel. Pers. Commun.* **2018**, *103*, 1791–1818. [[CrossRef](#)]
22. Wang, J.M.; Wang, Y.; Dai, X.M.; Bensaou, B. SDN-Based Multi-Class QoS Guarantee in Inter-Data Center Communications. *IEEE Trans. Cloud Comput.* **2019**, *7*, 116–128. [[CrossRef](#)]
23. Sun, L.; Tian, H.; Sun, Q.Y.; Shen, D.M.; Zhang, P. Traffic Allocation Scheme with Cooperation of WWAN and WPAN. *IEEE Commun. Lett.* **2010**, *14*, 551–553. [[CrossRef](#)]
24. Huang, Y.; Tian, H.; Zhang, J.; Qin, C. Rate Allocation Scheme for Machine-to-Machine Service Based on 3GPP in Heterogeneous Wireless Networks. *China Commun.* **2013**, *10*, 65–71. [[CrossRef](#)]
25. Zheng, J.; Li, J.D.; Liu, Q.; Shi, H.; Yang, X. On Minimizing Delay with Probabilistic Splitting of Traffic Flow in Heterogeneous Wireless Networks. *China Commun.* **2014**, *11*, 62–71. [[CrossRef](#)]
26. Goyal, S.; Tan, B.L.; Chincholi, A.; Elkourdi, T.; Demir, A. On the packet allocation of multi-band aggregation wireless networks. *Wirel. Netw.* **2018**, *24*, 2521–2537. [[CrossRef](#)]
27. Li, M.F.; Chen, H.L. Energy-Efficient Traffic Regulation and Scheduling for Video Streaming Services over LTE-A Networks. *IEEE Trans. Mob. Comput.* **2019**, *18*, 334–347. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).