Sustainable Virtual Network Function Placement and Traffic Routing for Green Mobile Edge Networks

Junbin Liang, Shaodong Huang, Yu Qiu, Lu Liu, *Member, IEEE*, Furqan Aziz, Min Chen, *Fellow, IEEE*

***Abstract*—Green Mobile Edge Networks (GMENs) are emerging networks that harvest green energy for powering mobile edge nodes, thereby reducing carbon dioxide emissions and energy costs. In GMENs, network service providers can flexibly place multiple virtual network functions (VNFs) that form a service function chain (SFC) in a specific order on geographically distributed edge nodes based on the level of harvested green energy, providing customized and sustainable network services for users. To meet the diversified availability requirements of users, backup SFCs need to be provided in addition to the primary SFC. These backup SFCs can be activated for providing uninterrupted services when the primary SFC is unavailable. However, due to the dynamic nature of wireless communication links, the uncertainty and unpredictability of green energy, and the limited resources available at edge nodes, optimizing the VNF placement and route traffic in real-time is challenging to minimize energy costs of all nodes and form expected SFCs with higher availability than user demand value. In this paper, the above problem is first formulated as an integer nonlinear programming and proven to be NP-hard. Then, it is discretized into a sequence of one-slot optimization problems to handle real-time changes in green energy and link availability. Finally, an online approximation strategy with a constant approximation ratio is proposed to solve the one-slot problems in polynomial time. This is the first study into online link availability-aware VNF placement and traffic routing problems in GMENs, motivated by sustainability concerns. The evaluation results indicate that the proposed scheme can ensure service availability while reducing the energy costs of all edge nodes and has achieved better performance when compared with other state-of-the-art methods.**

***Index Terms*—Mobile edge network, network function virtualization, service function chaining, placement, routing, green energy harvesting, availability.**

I. Introduction

A. Background and Motivation

The green mobile edge network is composed of mobile edge nodes that can harvest green energy (e.g., solar-integrated smart cars [1], buses [2], and drones [3], [4]), as well as a set of communication links between these nodes. Their aim is to provide users with low latency and sustainable green computing. The emergence of Network Function Virtualization (NFV) technology enables VNFs in the form of software instances to run on mobile edge nodes, achieving the same functions as middleboxes built on specialized hardware [5]. A user would request a series of VNFs, which form a service function chain (SFC) in a specific order to process traffic [6]. Meanwhile, each user would have a specific availability requirement for the SFC [7]. On the one hand, VNFs can be flexibly placed on different edge nodes according to the distribution of green energy, aiming to efficiently utilize green energy and reduce energy costs [8]. On the other hand, with increasing demands for network performance by 6G, e.g., availability and security, the energy consumption of future network service provisioning will continue to rise [9]. Consequently, driven by considerations of cost-effectiveness, high service quality, and sustainable network development, network service providers (NSPs) combine the two to provide tailored network services (e.g., high-precision mapping) for users. Fig.1 depicts a network scenario, where the network service provider places multiple VNFs based on the green energy harvested by the mobile edge nodes and plans traffic routing. In this paper, solar-integrated vehicles are regarded as mobile edge nodes that can harvest green energy.

However, mobile edge nodes are likely affected by ex-

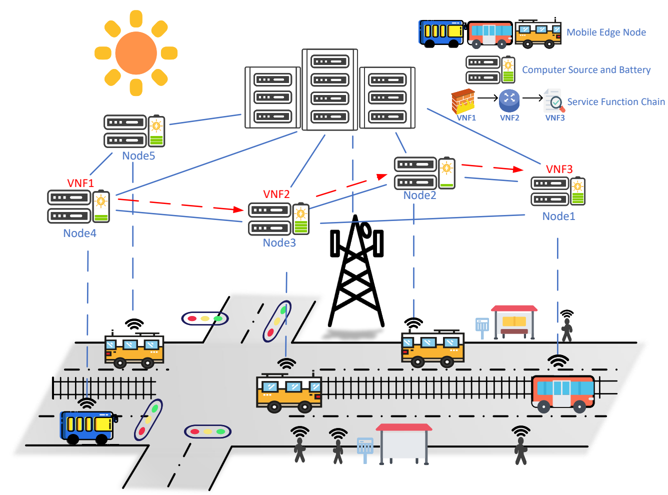


Fig. 1. Green Mobile Edge Networks

This research is funded by the National Natural Science Foundation of China (Grant No. 62362005) and China Scholarship Council (No. 202206665001).

Junbin Liang and Shaodong Huang are with the Guangxi Key Laboratory of Multimedia Communications and Network Technology, School of Computer, Electronics and Information, Guangxi University, Nanning, China. E-mail:huangsd9907@163.com.(Corresponding author: Shaodong Huang.)

Yu Qiu and Min Chen are with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, P. R. China. E-mail: csqiuyu@mail.scut.edu.cn. Lu Liu and Furqan Aziz are with the Department of Computing and Mathematic Sciences, University of Leicester, United Kingdom. E-mail: l.liu@leicester.ac.uk.

ternal factors and become unavailable. If the node becomes unavailable, its running VNF cannot provide network services [10]. In addition, mobile edge nodes use wireless links for communication. Wireless links are susceptible to interruptions caused by obstacles, weather, and distance between nodes, which can lead to data transmission failures for SFCs passing through the links [11]. The distributed placement of VNFs making up an SFC can balance load, battery energy and energy consumption of edge nodes, avoiding missing the chances of storing green energy due to batteries being fully charged. Thus, it enables efficient utilization of green energy, copes with the differences among nodes (e.g., computational resources and rate of harvesting green energy), and reduces network energy costs. However, due to the susceptibility of wireless communication links between mobile edge nodes to interruption and failure, decentralized placement of VNFs can lead to a decrease in SFC availability. If SFCs are frequently unavailable, virtual network services in the GMENs will be unsustainable because they fail to meet the users' application requirements. Therefore, NSPs need to make a trade-off between utilizing green energy and ensuring SFC availability.

Providing redundancy is a practical strategy for improving SFC availability. Multiple backup SFCs are provided until the availability threshold of the service is met, and they can continue to provide services when primal SFC fails [12]. However, due to limited communication range of mobile edge nodes and their dynamic changes in direction, speed, and acceleration, the probability of communication link interruption between nodes is not static but fluctuates with the distance between nodes. For instance, based on the wireless link evaluation formula [11], Figure 2 illustrates the time-varying curves depicting the positions of nodes 1, 2, and 3 in relation to the availability of the links between them. These nodes are shown in Figure 1. Nodes 1 and 2 are moving in opposite directions, while Nodes 1 and 3 are moving in the same direction, within the vehicle's communication range of 200 meters. On the other hand, due to weather changes and building obstruction, the distribution of green energy in time and space is uneven. The green energy harvested by each mobile edge node is uncertain and difficult to predict. If the energy consumption of the service on the node exceeds the harvested green energy, it is necessary to activate a backup power source (e.g., diesel) to ensure the stable operation of the service. Unfortunately, backup power has a higher energy cost compared to green power supply. Therefore, changes in network topology will make the previous SFC solution unable to always meet the availability requirements of network services and fully utilize green energy to reduce energy costs.

In addition, as mobile edge nodes are allowed to dynamically enter or leave the network, the number of nodes in the network also varies and is uncertain. This implies that the number of batteries, computational, and communication resources in the network keep changing with time. The above dynamic characteristics will couple current VNF placement and traffic routing decisions with future decisions. If the green energy of edge nodes that are about to leave the network is not utilized to save the green energy of nodes in future time slots, the remaining green energy of nodes may not be sufficient to provide network services in future time slots, which requires more backup power. Therefore, network service providers need to take into account the dynamic nature of the network to make optimal VNF placement and traffic routing joint decisions.

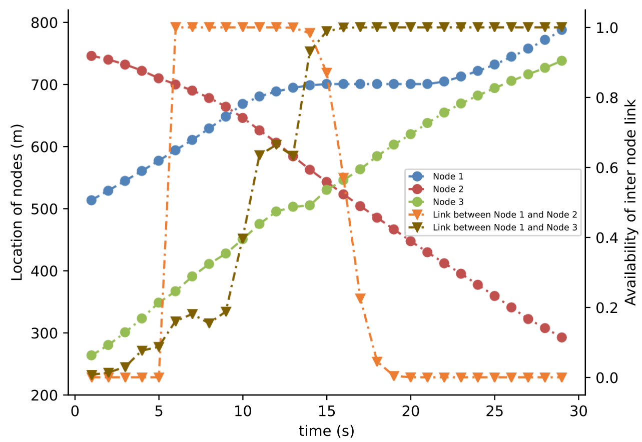


Fig. 2. The mobility of nodes and the availability of links between nodes

In this paper, the above problem is referred to as long-term green reliable SFC provisioning (LGRSP). Solving this problem will face the following challenges: (1) In GMENs, the computational resources, maximum battery capacity, green energy harvesting rate, and energy consumption per unit of computational resource of different nodes are different. This makes it difficult to determine placement locations of VNFs to meet the constraints of computational resources and battery capacity of the edge nodes; (2) In GMENs with limited communication resources, it's difficult to determine the routing path of the traffic to meet the order requirements of VNFs in SFC and the bandwidth resource constraints of links; (3) The decentralized placement of VNFs can harness green energy from different nodes to reduce energy costs. However, the nodes for placing VNFs of the SFC determine the starting, intermediate, and ending nodes of the route of the SFC. Moreover, the route significantly impacts the availability of the SFC composed of VNFs, as the wireless links constituting the route are prone to disruptions. Therefore, the VNF placement and route decisions are interdependent and there is a contradictory relationship between ensuring service availability and utilizing green energy, in addressing the problem aiming to minimize energy costs through green energy utilization while ensuring SFC availability in real-time. Furthermore, current and future decisions are coupling relationship in the time dimension, due to the dynamic and uncertain of green mobile edge networks; (4) Finally, the availability of nodes and links, as well as the availability requirements of users, all have differences. Therefore, it is very challenging to determine the number of backup SFCs, and their placement and routing decisions to meet the availability requirements of users in real-time.

B. literature review

1) Green Mobile Edge Networks: In recent years, many research efforts have been devoted to installing energy harvesting devices at edge nodes to harness renewable green energy from the environment onsite. The goal is to reduce the energy cost of network services and alleviate the sustainability issues caused by the explosive growth of network service demand [13]. Chen et al. [14] considered the problem of multi-user multi-task offloading and proposed two greedy scheduling algorithms, centralized and distributed, to maximize the benefit of receiving tasks. Ku et al. [15] and Ku et al. [16] further considered the fluctuating nature of green energy (e.g., solar) generation and proposed a two-stage algorithm and heuristic algorithm to find optimal task offloading decisions, respectively.

The above research considers geographically fixed edge servers. However, some research works have installed energy harvesting devices (e.g., solar panels) for mobile edge nodes. Sun et al. [3] used solar-powered UAVs to provide communication services to users on the ground. Sekander et al. [4] studied the amount of energy harvested by a multi-rotor UAV capable of harvesting solar and wind energy and derived a probabilistic model for UAV energy disruption. Considering the geographical uneven distribution of solar energy, Zhou et al. [1] proposed a greedy algorithm to solve the driving route of a solar-powered self-driving car to maximize the harvested solar energy.

2) Service function chains availability: Once a node or a link in a SFC fails, the SFC fails to operate and hence the service has to be stopped. This is unacceptable for emerging IoT applications, which have stringent requirements on network performance [17]. Ali et al. [18] deployed a backup for each VNF to guarantee service availability. However, the length of the protected SFC increases as the data flow needs to go through backup instances, which results in longer end-to-end latency of the service. Qu et al. [19] proposed heuristic algorithms to solve VNF deployment and routing decisions that minimize link utilization and satisfy the latency and reliability requirements. To cope with SFC interruption due to node or link failure, Yang Song et al. [12] placed additional K groups of backup SFCs for each service and plans their data routing paths. Mai et al. [20] further proposed a heuristic algorithm that reduces energy consumption by sharing VNF.

These research efforts determine the probability of failure of a VNF instance based on the average of previous historical data and assume that the failure rate of VNF instances is static. However, many failures are highly correlated with one another and failure spikes may happen at any time [21]. Shang et al. [22] proposed a dynamically added backup VNFS scheme to meet the reliability requirements of services in real time and at the lowest cost. Qiu et al. [7] further considered the reliability of VNF instances changes dynamically as the load changes, and proposed an approximation algorithm to dynamically deploy VNFs and backup. However, the above research considers the provision of network services at fixed edge nodes. Balazs et al. [23] further considered the reliability problem of wireless communication links for mobile edge nodes (i.e., vehicles), which requires that the probability of a mobile node being covered by a radio must satisfy a threshold as a way of preventing the interruption of its communication link. In order to fully utilize the resources of fog (e.g., vehicles) and edge nodes, Jorge et al. [24] proposed heuristic algorithms to solve the VNF placement and routing decisions to minimize costs and to satisfy the reliability of services.

3) Green reliable service function chain provisioning: Sameer et al. [25] proposed a green VNF deployment framework based on renewable energy harvested by geographically dispersed servers, which is realized by backing up VNF instances and server energy deficit warning technique. When the green energy harvested by the server is insufficient to run the VNF, the backup VNF which is running on an edge node powered by the grid is used instead of the VNF to continue providing services. However, they focused on providing an energy-efficient VNF deployment framework, ignoring the availability issues of links and SFC. Shang et al. [21] used renewable green energy and a limited budget to power the backup VNF. However, this study is based on centralized green energy harvesting and overlooks the impact of distributed green energy on reliable VNF placement and traffic routing decisions.

TABLE I
COMPARISON OF OUR WORK WITH EXISTING RELATED STUDIES

References	**VNF placement**	**Traffic routing**	**Green energy**	**Availability of SFC**
[1], [3], [4], [13]–[16]	✗	✗	✓	✗
[7], [18], [22], [23]	✓	✗	✗	✓
[12], [19], [20], [24]	✓	✓	✗	✓
[25]	✓	✗	✓	✗
[21]	✗	✗	✓	✓
Our work	✓	✓	✓	✓

4) Novelty and feasibility: The primary novel aspect is the problem we consider in this paper. The problem is a joint optimization problem of VNF placement and traffic routing. It simultaneously considers two factors, i.e., (i) the availability of wireless links and the green energy harvested by nodes with limited battery capacity and computational resources, and (ii) needs to make a balance between the availability of SFC and the utilization of green energy. We compare our work with the existing related works in table I.

Our research work is feasible in the near future, as it promotes the sustainable development of green and reliable mobile edge networks. At present, there is already work to place VNFs on mobile edge nodes, such as vehicles [23], [26]–[28], drones [29], [30], and robots [24], to provide network services. In addition, solar powered vehicles will also be gradually promoted in the future. For instance, Sono Motor [2] had developed solar-powered smart vehicles for real-world applications, specifically passenger cars, refrigerated trucks, and buses. There is research work [1] that studied the movement trajectory of solar-powered vehicles to deal with unevenly distributed solar energy. Furthermore, there is the existence of multiple pathways between vehicle nodes [31], [32]. These pathways can serve as backup routing for data transmission.

C. Contributions

To address the above mentioned challenges of the LGRSP problem, in this paper we first formulate the problem as an integer nonlinear programming and prove that it is NP-hard by reducing it to the Lowest cost Generalized Assignment Problem (LGAP) [7]. Next, the LGRSP problem is discretized into a series of one-slot optimization problems, where VNF placement and traffic routing strategies are adaptively adjusted according to the network topology in each slot to handle actual failure rates and utilize gathered green energy. Finally, we propose a two-stage online scheme to solve the one-slot optimization problem. In the first stage, the scheme prioritizes the comprehensive utilization of green energy from diverse nodes, as it aligns with the objective of minimizing energy consumption costs. So, it places VNFs of SFCs in a decentralized manner on some nodes with maximal gains in green energy. However, the availability requirement of SFCs may not be met consistently, due to the dynamic of green mobile edge networks. Therefore, in the second phase, the scheme enhances the availability of SFCs by providing backup SFCs oriented to mitigate the impact of link failures on service availability. It places the VNFs of the backup SFCs in the same node as much as possible, which has the most abundant green energy. In both stages, the placement nodes of the VNFs are determined first, followed by the shortest paths between the nodes are used as routing paths for the traffic, decoupling VNF placement and traffic route decisions.

The main contributions of this paper are as follows:

- To the best of our knowledge, we are the first to study the online link availability-aware VNF placement and traffic routing problem in GMENs, which requires a trade-off between utilizing green energy and ensuring SFC availability. Moreover, this problem also considers the volatility of wireless link availability and the uncertainty of green power, which is formulated as an integer nonlinear programming problem.
- In order to better capture the time-varying availability of wireless links and handle the uncertainty of green energy, LGRSP is discretized into a series of one-slot optimization problems, both of which have been proven to be NP-hard.
- We propose a two-stage online scheme with a constant approximation ratio, that can solve one-slot problems in polynomial time.

The rest of this paper is organized as follows. Section II and Section III introduces system model and problem statement, respectively. Next, in section IV, an online scheme is proposed to solve LGRSP. Section V evaluates the performance of our solution. Finally, the conclusion of our research work is provided in Section VI.

II. SYSTEM MODEL

A green mobile edge network is composed of mobile edge nodes and communication links between the nodes. Mobile edge nodes in this paper are solar-integrated smart vehicles [33], [34], which are equipped with computing devices, antennas, cellular interfaces, solar panels, and a Global Positioning System. They have computing power, storage, communication, and green energy harvesting capabilities. The research scenario in this paper is a straight two-way multi-lane road in an urban environment [11]. The continuous time interval $T = \{t_1, t_2, ..., t_{|T|}\}$ of the system is divided into several equal time slots each with a time duration Δt. The road-side network facilities (e.g., fixed edge nodes with base stations) act as control centers and implement centralized network service management using a combined SDN/NFV network architecture [27]. It is assumed that multiple nodes that have been covered by the roadside network facilities and their wireless communication links with the roadside facilities are reliable and able to receive/send data packets, such as SDN control data, normally.

A. Substrate network

We consider a GMENs to be an undirected weighted graph $G(t) = (N(t), E(t))$, where $N(t) = \{n_1, n_2, \ldots, n_{|N(t)|}\}$ represents the set of mobile edge nodes, and $E(t) = \{e_{i,j} | n_i, n_j \in N(t)\}$ is the set of wireless communication links between the nodes. $e_{i,j}$ is the wireless communication link between nodes n_i and n_j at time t. The communication range, bandwidth and computational resources of mobile edge nodes are limited. The maximum communication range of a node is assumed to be the constant D_{comm}. A link $e_{i,j}$ exists if the geographic distance between nodes n_i and n_j is less than or equal to D_{comm}. For simplicity, we assume that the total bandwidth of each link $e_{i,j} \in E(t)$ is the constant B_e, and each node $u \in N(t)$ has the computational resource cap_u.

B. Service model

The network service provider embeds SFCs for each user request to provide customized network services. These SFCs consist of multiple VNFs in a sequential order. The set of all requested SFCs is represented as S, and the set of VNFs for each requested SFC $s \in S$ is denoted by $F_s = \{f_s^1, f_s^2, \ldots, f_s^{H_s}\}$. Here, H_s represents the total number of VNFs in a SFC $s \in S$, and f_s^i denotes the ith VNF of SFC $s \in S$, which needs to occupy a certain amount of computational resources $c_{f_s^i}$ for its operation. The user's data is transferred among the H_s instances in the order of the VNFs in the F_s, i.e., $f_s^1 \rightarrow f_s^2 \rightarrow \ldots \rightarrow f_s^{H_s}$. Additionally, we use B_s and R_s to represent the bandwidth requirement and the availability requirement of requested SFC $s \in S$, respectively.

C. Availability model

The availability of a system is the fraction of time that the system is operational during the entire service time. The availability A_i of a network component i can be calculated as [12]:

$$A_i = \frac{MTTF}{MTTF + MTTR} \tag{1}$$

where MTTF represents Mean Time To Failure and MTTR denotes Mean Time To Repair. In this paper, the availability of a mobile edge node is defined as the product of the availability of all components associated with computation and communication (e.g., processors, antennas). The availability of devices can be obtained by accessing previous detailed logs that record every hardware component failure and maintenance event over the lifetime of the device.

Vehicles as mobile edge nodes are highly mobile and their communication range is limited. Therefore, during the movement of two nodes, if the relative distance between the nodes is greater than the communication range, the wireless communication link connecting them is unavailable. Compared to the availability of nodes, the availability of wireless links varies in real time and is unstable. In this paper, the wireless link model proposed by Zhang et al. [11] is used to measure the availability of links between vehicle nodes. For completeness of presentation, some of the important equations are revisited in this paper. The availability of link is the probability that the two vehicles continuously keep available within the communication range over a specified time interval. So, the availability of the link $e_{i,j}$ in time slot t can be obtained as follows:

$$A_{e_{i,j}}(t) = P\left\{D\left(u_i, u_j\right)_t < D^r_{i,j}(t)\right\} \tag{2}$$

where $D(u_i, u_j)_t$ is a variable, which represents the change in the relative distance between nodes u_i and u_j in the time slot $t \in T$. $D^r_{i,j}(t)$ is a constant, which represents the remaining communication distance between nodes u_i and u_j at time slot t according to the different cases, which can be classified as follow: $Case1$: Two nodes are moving on a road in same directions, so that their remaining communicable distance in the worst case is

$$D^r_{i,j}(t) = D_{comm} - d_{ij}(t) \tag{3}$$

$Case2$: Two nodes are moving on a road in the opposite directions,

$$D^r_{i,j}(t) = \begin{cases} D_{comm} - d_{ij}(t), \\ \textit{two nodes are driving away from each other} \\ D_{comm} + d_{ij}(t), \\ \textit{two nodes are driving towards each other} \end{cases} \tag{4}$$

where $d_{ij}(t)$ is the Euclidean distance of vehicle nodes u_i, u_j at the beginning of the time slot t. Suppose an unprotected SFC places VNFs on w nodes $n_1, n_2, \ldots, n_w$, and traverses m links $l_1, l_2, \ldots, l_m$. Whenever one of the nodes or links becomes unavailable, the SFC stops providing services. Therefore, the availability of the unprotected SFC is calculated as [12],

$$A^s_{upro}(t) = \prod_{i=1}^{w} A_{n_i} \prod_{j=1}^{m} A_{l_j}(t), n_i \in N(t), l_j \in E(t) \tag{5}$$

where $\prod_{i=1}^{w} A_{n_i}$ denotes the availability of all the used nodes and $\prod_{j=1}^{m} A_{l_j}(t)$ denotes the availability of all the traversed links at time slot t. In this paper, we adopt backup SFCs to protect the requested SFCs, enhancing the availability of them. For the requested SFC s, we consider it possesses $p_s(t)$ disjoint unprotected SFCs at time slot t, i.e., these SFCs do not place VNFs on the same node and do not pass through the same link. We term the first unprotected SFC in the $p_s(t)$ SFCs as the primary SFC. When the primary SFC is unavailable, $p_s(t) - 1$ backup SFCs are activated to continue serving users. To avoid too many backups taking up computational and communication resources, we limit the maximum $p_s(t)$ value of each requested SFC s to p_{max}. The availability of requested SFC $s \in S$ at t time slot can be calculated as [12]:

$$A^s_{pro}(t) = 1 - \prod_{i=1}^{p_s(t)} \left(1 - A^s_{upro}(t)\right), 1 \le p_s(t) \le p_{max} \tag{6}$$

D. Energy Model

The energy consumption of a mobile edge nodes $u \in N(t)$ for network services mainly includes two aspects. The first one is the energy requirement for basic operations and communication pb_u. This includes, for example, the energy consumption of the operating system and the signal transmitter, which is used to maintain its own basic operation [20]. The second one is the computational energy demand for network services. The energy consumption of an edge node u at full computational load is denoted as pm_u. Therefore, the total network energy consumption $CE_u(t)$ of an edge node u at time slot $t \in T$ is defined as:

$$\begin{aligned} &CE_u(t) = pb_u + \\ &(pm_u - pb_u) \cdot \frac{\sum\limits_{s \in S, 1 \le j \le p_s(t), 1 \le h \le H_s} x^u_{s,j,h}(t) \cdot c_{f^h_s}}{cap_u} \end{aligned} \tag{7}$$

where the binary variable $x^u_{s,j,h}(t)$ represents whether the $h \in [1, H_s]$th VNF of the $j \in [1, p_s(t)]$th disjoint unprotected SFC of requested SFC $s \in S$ is placed on the node $u \in N(t)$ at time slot $t \in T$.

The energy harvested by the solar panels can either power the edge nodes directly or charge the batteries. When the green energy is low, the batteries take over to keep the edge nodes running stably. Although there are models for harvesting green energy, it is difficult to accurately predict and evaluate the harvested energy during node movement. The reason is that the process of harvesting green energy by mobile nodes is easily blocked by local buildings, and the situation at each geographical location is unknown and uncertain. To model the uncertainty of green energy, we assume that the green energy $HE_u(t)$ harvested by node $u \in N(t)$ at time slot $t \in T$. $HE_u(t)$ is obtained only after the placement and routing decisions are made. In addition, $BE_u(t) \in [0, BE_{max}]$ is denoted as the battery energy of node $u \in N(t)$ at the beginning of the time slot $t \in T$, where BE_{max} refers to the capacity of the battery.

In reality, the vehicle can obtain a stable amount of energy from a gas station. Therefore, we assume that when the battery energy cannot satisfy the energy consumption of the edge node, backup power supply (e.g., diesel generator) can always be activated to keep the node running. However, the energy cost of the backup power will be greater than the energy cost of the green energy [35]. Assume that the cost factor of using green energy and using backup power at the node are α and $\beta (0 < \alpha \leq \beta)$, respectively. There are two possible cases: $Case1$: When $BE_u(t) < CE_u(t)$, the backup power supply will be activated and output the energy of $CE_u(t) - BE_u(t)$ units. Then, the battery energy $BE_u(t+1)$ in the next time slot is

$$BE_u(t+1) = \min(BE_{max}, HE_u(t)) \tag{8}$$

$Case2$: When $BE_u(t) \geq CE_u(t)$, the battery energy is sufficient to meet the energy consumption of the edge nodes. According to the harvested green energy $HE_u(t)$ and energy demand $CE_u(t)$, the battery is charged and discharged accordingly:

- If $HE_u(t) \geq CE_u(t)$, the residual energy will be stored in the battery until it reaches maximum capacity BE_{max}.

$$\begin{aligned} &BE_u(t+1) \\ &= \min(BE_{max}, BE_u(t) + HE_u(t) - CE_u(t)) \end{aligned} \tag{9}$$

- If $HE_u(t) < CE_u(t)$, then the battery has to be discharged to cover the energy deficit.

$$BE_u(t+1) = BE_u(t) + HE_u(t) - CE_u(t) \tag{10}$$

Therefore, in time slot $t \in T$, the energy cost of edge node $u \in N(t)$ is

$$Cost_u(t) = \begin{cases} BE_u(t) \cdot \alpha + (CE_u(t) - BE_u(t)) \cdot \beta, \\ \quad BE_u(t) \leq CE_u(t) \\ CE_u(t) \cdot \alpha, BE_u(t) > CE_u(t) \end{cases} \tag{11}$$

III. PROBLEM STATEMENT

A. Problem Definition

Definition 1. Given a green mobile edge network $G = (N, E)$ with limited communication, computational resources and node battery capacity, the one-slot green reliable SFC provisioning (OGRSP) problem is to place VNFs to the mobile edge nodes based on requested SFCs $s \in S$ and the green energy harvested by the nodes, and to route the traffic between these VNFs, such that: (1) The computational resource constraints of mobile edge nodes are not violated; (2) The battery capacity constraint of the node is not violated; (3) The bandwidth constraints of the link are not violated; (4) The data forwarding order between VNFs in SFC is not violated; (5) The availability of the requested SFC is not less than R_s; (6) The energy cost of all mobile edge nodes is minimized.

Definition 2. Given a green mobile edge network $G(t) = (N(t), E(t))$ with limited communication, computational resources and node battery capacity, the number of node, the geographic location of each node and the harvested green energy varies over any time slot $t \in T$. The Long-term Green Reliable SFC Provisioning (LGRSP) problem is to minimize long-term energy costs of all nodes, without any future information such as green energy, the number of nodes, the geographic location of each node, and outage probability of links, while making VNF placement and traffic routing decisions at the beginning of the time slot t in the time span T, so that the availability requirements of each requested SFC are satisfied in real-time.

B. NP-Hardness

In green mobile edge networks, the OGRSP and LGRSP problems are NP-hard, which is proved below. We prove that the OGRSP problem is a NP-hard problem by a reduction from a well-known NP-hard problem - the Least Cost Generalized Allocation Problem (LGAP). The definition of the LGAP problem is as follows, given a set of items $Item = (a_1, a_2, \ldots, a_n)$ with size $size_i$ and a set of bins $Bin = (b_1, b_2, \ldots, b_m)$ with capacity cap_{b_i}. If an item a_i is placed in the box b_j ($1 \leq i \leq n$ and $1 \leq j \leq m$), it incurs a cost $cost_{i,j} > 0$. The problem aims to minimize costs, pack as many items as possible into boxes, and ensure that the total size of items in each box does not exceed its capacity [36].

We show that the LGAP problem can be reduced to a special case Q of the OGRSP problem. Q is composed of $G = (N, E)$ and a set of requests S, which is a problem that ignores the availability requirements of the services, the basic energy cost of the nodes, the green energy and batteries, and the traffic routing. Specifically, each edge node $u \in N$ has computational resource cap_u, its base and full-load energy consumption are pb_u and pm_u, respectively, and the SFC of each request $s \in S$ consists of just one VNF, which requires computational resource of f_s. The energy cost of the edge node $u \in N$ to run the instance of the VNF of the request $s \in S$ is $cost_{u,s} = (pm_u - pb_u)\frac{f_s}{cap_u}$. Thus, the OGRSP problem aims to satisfy the computational resource constraints of mobile edge nodes to receive as many requests as possible with minimum energy cost.

Obviously, if there is a solution to the special case Q of OGRSP problem, there is a solution to the LGAP problem. The LGAP problem is known to be NP-hard, so the OGRSP problem is also NP-hard. Assuming that the time span T only contains one time slot, i.e., $T = 1$, the LGRSP problem will be simplified as an OGRSP problem. Therefore, the LGRSP problem is also NP-hard.

C. Integer nonlinear programming model

In this part, relevant symbols are first listed in the TABLE II. Then, LGRSP problem is formulated as an integer nonlinear program. Decision variables include $x^u_{s,j,h}(t), y^{e_{u,v}}_{s,j,h}(t), p_s(t)$.

Objective:

$$minimize \sum_{t \in T, u \in N(t)} Cost_u(t) \tag{12}$$

Subject to:

$$0 \leq BE_u(t) \leq BE_{max}, \forall t \in T, u \in N(t) \tag{13}$$

$$\sum_{u \in N(t)} x^u_{s,j,h}(t) = 1, \forall t \in T, s \in S, 1 \leq j \leq p_s(t), 1 \leq h \leq H_s \tag{14}$$

$$\begin{aligned} &\sum_{e_{u,v} \in E(t)} y^{e_{u,v}}_{s,j,h}(t) = 1, \forall t \in T, s \in S, 1 \leq j \leq p_s(t), \\ &1 \leq h \leq H_s - 1, u \in N(t) : x^u_{s,j,h}(t) = 1 \&\& x^u_{s,j,h+1}(t) \neq 1 \end{aligned} \tag{15}$$

TABLE II
NOTATIONS

Symbol	Description	Symbol	Description
$T=\{t_1,t_2,...,t_{\lvert T\rvert}\}$	Time set	$G(t)=(N(t),E(t))$	The undirected graph corresponding to GMENs network in time slot t, where $N(t)$ is the set of nodes and $E(t)$ is the set of edges.
$s\in S$	The set of SFC required by users	u	The mobile edge node $u\in N(t)$
$F_s=\left\{f_s^1,f_s^2,\ldots,f_s^{H_s}\right\}$	The set of VNFs for each requested SFC	$e_{u,v}\in E(t)$	The link between node u and v at time slot t
H_s	The length of SFC $s\in S$	cap_u	The computational resource of node $u\in N(t)$
f_s^i	The ith VNF of $s\in S$	D_{comm}	The range of node's communication
$c_{f_s^i}$	The computational resource of VNF f_s^i	B_e	The bandwidth resource of each link $e_{i,j}(t)\in E(t)$
B_s	The bandwidth requirement of SFC $s\in S$	A_n	The availability of node $n\in N(t)$
R_s	The availability requirement of SFC $s\in S$	$A_{e_{i,j}}(t)$	The availability of link $e_{i,j}\in E(t)$
$x_{s,j,h}^u(t)$	Bool variables, which identify whether VNF f_s^h of jth SFC for $s\in S$ is placed on node u at time slot $t\in T$	$y_{s,j,h}^{e_{u,v}}(t)$	Bool variables, which identify whether the traffic between VNF f_s^h and f_s^{h+1} of jth SFC for $s\in S$ traverse link $e_{u,v}(t)$ at time slot $t\in T$
$p_s(t)$	p_s disjoint unprotected SFCs for $s\in S$ at $t\in T$	α,β	The cost coefficients of green energy and backup power, respectively
$CE_u(t)$	The total energy consumption of an edge node u at time slot t	$BE_u(t)$	The battery energy of node u at time slot t
$HE_u(t)$	The green energy harvested by node u at time slot t	$Cost_u(t)$	The energy cost of edge node u in time slot t
pb_u	The energy requirement of node u for basic operations	pm_u	The energy consumption of node u at full computational load

$$\sum_{e_{u,v}\in E(t)} y_{s,j,h}^{e_{u,v}}(t)=1,\forall t\in T,s\in S,1\le j\le p_s(t),$$
$$1\le h\le H_s-1,v\in N(t):x_{s,j,h}^v(t)\ne 1\&\&x_{s,j,h+1}^v(t)=1 \tag{16}$$

$$\begin{aligned}\sum_{e_{u,v}\in E(t)} y_{s,j,h}^{e_{u,v}}(t)=\sum_{e_{v,w}\in E(t)} y_{s,j,h}^{e_{v,w}}(t),\forall t\in T,s\in S,\\ 1\le j\le p_s(t),1\le h\le H_s-1,\\ v\in N(t):x_{s,j,h}^v(t)\ne 1\&\&x_{s,j,h+1}^v(t)\ne 1\end{aligned} \tag{17}$$

$$\sum_{s\in S,1\le j\le p_s(t),1\le h\le H_s} x_{s,j,h}^u(t)\cdot c_{f_s^h}\le cap_u,\forall t\in T,u\in N(t) \tag{18}$$

$$\sum_{s\in S,1\le j\le p_s(t),1\le h\le H_s-1} y_{s,j,h}^{e_{u,v}}(t)\cdot B_s\le B_e,\forall t\in T,e_{u,v}\in E(t) \tag{19}$$

$$\sum_{1\le j\le p_s(t)} x_{s,j,h}^u(t)\le 1,\ \forall t\in T,s\in S,u\in N(t),1\le h\le H_s \tag{20}$$

$$\begin{aligned}\sum_{1\le j\le p_s(t)} y_{s,j,h}^{e_{u,v}}(t)\le 1,\forall t\in T,\\ s\in S,e_{u,v}\in E(t),1\le h\le H_s-1\end{aligned} \tag{21}$$

$$A_{pro}^s(t)\ge R_s\ ,\forall t\in T,s\in S \tag{22}$$

$$1\le p_s(t)\le p_{max},\forall t\in T,s\in S \tag{23}$$

$$\begin{aligned}x_{s,j,h}^u(t),y_{s,j,h}^{e_{u,v}}(t)\in\{0,1\},\forall t\in T,s\in S,u,v\in N(t),\\ e_{u,v}\in E(t),1\le j\le p_s(t),1\le h\le H_s\end{aligned} \tag{24}$$

Eq. (13) is a battery capacity constraint that requires each node's battery level to not exceed its maximum capacity. This battery is used to store the harvested green energy. That is to say, the green energy that each node can use is limited. Eq. (14) is placement constraint, which requires each VNF can only be placed on one node. Eq. (15)(16)(17) is a flow conservation constraint for traffic routing. Eq. (20) (21) are the VNF disjoint constraint and the link disjoint constraint, respectively, which ensure that the $p_s(t)$ SFC of protected SFC s are not intersecting with each other. Note that it can be easily extended to other mathematical models that include latency constraints and migration decisions.

IV. OTGRSP APPROXIMATION SCHEME

In practice, NSPs have the following needs for an algorithm to solve the LGRSP problem. 1) the algorithm should have low running time to adapt to the highly dynamic nature of GMENs and to meet the real-time requirements of user requests. 2) the algorithm should be explainable and have performance bounds, because industrial applications need to ensure the stability and predictability of the execution results. However, Lyapunov methods are characterized by a long execution time, in large-scale GMENs, because they require a large number of iterative steps to reach a satisfactory solution, i.e., slow convergence. Moreover, they are more suitable for certain and

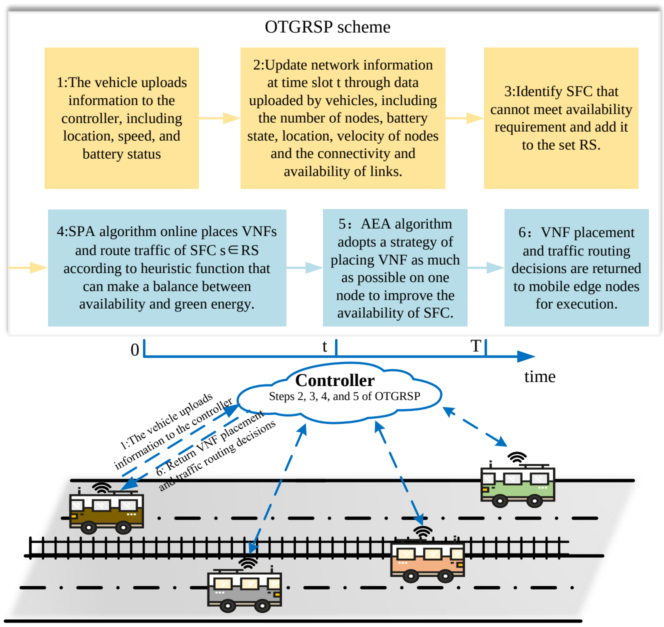


Fig. 3. The process of OTGRSP

known systems, and for assessing the stability of the system. Therefore, we propose an Online Two-stage Green Reliable SFC Provisioning (OTGRSP) scheme to solve the LGRSP problem. It is an online approximation scheme with provable upper and lower bounds on performance and has polynomial time complexity.

A. OTGRSP scheme overview

OTGRSP consists of two approximation algorithms: the SFC Provisioning Algorithm (SPA) and the Availability Enhanced Algorithm (AEA). The processing pipeline of the OTGRSP is shown in Fig. 3. Intuitively, the smaller the routing path length of the SFC, the higher the SFC availability will be, as long as the availability of all links is the same. However, this also reduces the possibility of placing VNFs to nodes with sufficient green energy. On the contrary, while decentralized placement of multiple VNFs of SFCs to different nodes can increase the green energy utilization and reduce the energy cost, longer paths can lead to a decrease in the availability of the requested SFC. Therefore, the SPA algorithm increases the chances of utilizing green energy by placing multiple VNFs that make up the primal SFC at different nodes and using the shortest path to route data between these nodes. Then, to cope with dynamic changes in green energy and link availability, the AEA algorithm adaptively adjusts VNF placement and traffic routing schemes, as well as improves the availability of the SFC by placing the backup SFC on a node as much as possible. The pseudocode for the solution process of SPA, AEA, and OTGRSP are shown in Algorithm 1, Algorithm 2, and Algorithm 3, respectively.

B. SFC Provisioning Algorithm (SPA)

For requested SFC $s \in S$, SPA firstly takes a node that meets the computational demand of VNF f_s^1 and has the maximal green energy as the source point o. If $|F_s| \geq 2$, nodes within $|F_s| - 1$ hop range centered at the source point o are included in the candidate set $Cand = \{n | dis\,(o,n) \leq |F_s - 1|\,, o, n \in N\}$, which is shown in lines 5-9. Then, calculate the green energy benefits of placing VNF f_s^2 to each node $n \in Cand$ in the candidate set by $bft_n = \frac{BE_n(t) \geq (pm_n - pb_n)\frac{c_{f_s^2}}{cap_n} ? (pm_n - pb_n)\frac{c_{f_s^2}}{cap_n} : BE_n(t)}{(1-R_{o,n})\cdot(p_{max}-1)}$, which is a heuristic function with ternary operator. $1 - R_{o,n}$ is the links unavailability of the shortest feasible path from the source point o to the node n, which is determined by the Dijkstra algorithm and its bandwidth should be greater than or equal to the bandwidth requirement of the requested SFC. $BE_n(t) \geq (pm_n - pb_n)\frac{c_{f_s^2}}{cap_n} ? (pm_n - pb_n)\frac{c_{f_s^2}}{cap_n} : BE_n(t)$ represents the green energy provided by node n for VNF f_s^2. If the value of $1 - R_{o,n}$ is higher, it indicates a higher possibility of placing additional $p_{max} - 1$ VNFs to improve the availability of SFC, which will consume energy, computational resource. So, We use $\frac{1}{(1-R_{o,n})\cdot(p_{max}-1)}$ as the penalty weight for placing VNF at the node $n \in Cand$.

Next, select the node o' with the highest green energy benefit in the candidate set, is shown in line 10. The computational resource of the node o' meets the requirements to place VNF f_s^2. If there are no nodes that meet the requirement, then the time slot is not suitable for placing the request on the mobile edge node and releasing the VNF instance previously placed for the SFC. Intuitively, this means that the computational or communication resources of the nodes within the range cannot support the VNF placement and routing. We will add it to $RS(t+1)$ and attempt to serve it in the next time slot, which is shown in lines 11-14. After placing f_s^2, SPA updates the computational resources of o and o', as well as the communication resources of the o to o' routing path, and record o' as the new source point o, which is shown in lines 17-18. In the above way, iteratively select nodes to place $f_s^i\,(3 \leq i \leq |F_s|)$. The difference is that the distance between the nodes included in the candidate set and the source point is $|F_s| - i$, which limits the distance between f_s^{i-1} and f_s^i, controlling the possible maximum path length of SFC. Moreover, candidate nodes of f_s^i are not nodes that have been selected to place VNFs, which is shown in line 6, so SPA avoids that VNFs of the SFC are centrally placed on a single node and result in overloading the node.

C. SPA Algorithm analysis

In this section, we will prove the upper bound of the total energy cost of the network for processing all requests within a specific time slot. Subsequently, the approximation ratio and time complexity of Algorithm 1 were analyzed and calculated. To simplify the proof, it is assumed that all requested VNFs can be placed to edge nodes within that specific time slot, and there is a link that can route user data.

Lemma 1. In green mobile edge networks, the upper bound of the network energy cost for algorithm SPA is $C = \left(C_{base} + \lambda_{max} \cdot \sum_{s \in S} \sum_{i \in F_s} c_{f_s^i}\right) \cdot \beta$, where $C_{base} = \sum_{n \in N} pb_n$ and $\lambda_{max} = \max_{n \in N} \frac{(pm_n - pb_n)}{c_n}$ are constant. pb_n and pm_n are

Algorithm 1: SPA

Input: $G(t) = (N(t), E(t))$, S
Output: $D1[i][j][h]$, $D2[i][j][h]$, $RS(t+1)$
1 **foreach** $s \in S$ **do**
2 $\quad o \leftarrow n = \max\limits_{n \in \{u | u \in N(t), cap_u(t) \geq c_{f_s^1}\}} BE_n(t)$;
3 $\quad D1[i][j][h] \leftarrow o; path_s \leftarrow o$;
4 $\quad$ **foreach** $f_s^i \in F_s \backslash f_s^1$ **do**
5 $\quad\quad Cand \leftarrow \emptyset$;
6 $\quad\quad$ **foreach** $n \in \{x | \ x \in N(t), x \notin path_s\}$ **do**
7 $\quad\quad\quad$ **if** $cap_n \geq c_{f_s^i}$ *and* $shortdistance(o, n) \leq |F_s| - i$ **then**
8 $\quad\quad\quad\quad Cand \leftarrow Cand \cup n$
9 $\quad\quad$ **end**
10 $\quad\quad o' \leftarrow$ the node $w \in Cand$ with the highest bft_w value.
11 $\quad\quad$ **if** *not* o' **then**
12 $\quad\quad\quad RS\,(t+1) = RS\,(t+1) \cup s$;
13 $\quad\quad\quad$ Update resources for nodes and links;
14 $\quad\quad\quad$ **break**;
15 $\quad\quad$ Establish shortest path $\Psi_{o,o'}$ between o and o' with Dijkstra Algorithm, according to the traffic load of the SFC s and the communication resources of the network;
16 $\quad\quad D1[s][1][i] \leftarrow o'$, $D2[s][1][i] \leftarrow \Psi_{o,o'}$;
17 $\quad\quad$ Update resources for nodes and links;
18 $\quad\quad path_s = \ path_s + \ \Psi_{o,o'} + o'$; $o \leftarrow o'$
19 $\quad$ **end**
20 **end**

the basic and full load energy consumption of node n, $c_{f_s^i}$ is CPU unit requirement of VNF f_s^i, β is the unit energy cost of backup power supply, N is the set of edge nodes at a time slot.

proof. The energy consumption of a network is generated by two parts: the basic operation of nodes and the processing of network functions. For the former, it is a constant $C_{base} = \sum\limits_{n \in N} pb_n$. The latter can be calculated based on the energy consumption required to process each requested SFC. The energy consumption of each VNF for processing requested SFC $s \in S$ is C_i. Therefore, the energy consumption of each requested SFC $s \in S$ is C_s.

$$
\begin{aligned}
C_s = \sum_{i \in F_s} C_i =& \sum_{i \in F_s} \frac{c_{f_s^i}}{c_n} \cdot (pm_n - pb_n) \\
& \leq \sum_{i \in F_s} c_{f_s^i} \cdot \lambda_{max} \quad , \lambda_{max} = \max_{n \in N} \frac{pm_n - pb_n}{c_n}
\end{aligned} \tag{25}
$$

Lemma 2. Given a green mobile edge network that has limited battery capacity, computational, and communication resources, and needs to handle requests S in time slot $t \in T$, there is an approximation algorithm (SPA) that can place VNFs and route traffic, aiming to minimize the energy cost of all nodes. The upper and lower bound of approximate ratio of SPA are $\frac{C_{base} + \lambda_{max}\theta}{\theta \cdot \lambda_{min}} \cdot \frac{\beta}{\alpha}$ and $\frac{C_{base} \cdot \alpha}{C_{base} \cdot \alpha + \lambda_{max} \cdot \beta \cdot \theta}$, respectively, and the time complexity is $O\left(|S| \cdot H_s \cdot |N(t)|\right)$, where $C_{base} = \sum\limits_{n \in N} pb_n$, $\lambda_{max} = \max\limits_{n \in N} \frac{(pm_n - pb_n)}{c_n}$, $\lambda_{min} = \min\limits_{n \in N} \frac{(pm_n - pb_n)}{c_n}$, $\theta = \sum\limits_{s \in S} \sum\limits_{i \in F_s} c_{f_s^i}$, $N = |N(t)|$ is the number of nodes at t time slot.

Proof. Assuming C^* is the optimal solution to the problem, C is the cost of OTGRSP scheme.

$$
\begin{aligned}
\frac{C}{C^*} &\leq \frac{\left(\sum\limits_{n \in N} pb_n + \lambda_{max} \cdot \sum\limits_{s \in S, i \in F_s} c_{f_s^i}\right) \cdot \beta}{\left(\sum\limits_{n \in N} pb_n + \sum\limits_{s \in S, i \in F_s} c_{f_s^i} \cdot \lambda_{min}\right) \cdot \alpha} \\
&\leq \frac{C_{base} + \lambda_{max} \cdot \theta}{\theta \cdot \lambda_{min}} \cdot \frac{\beta}{\alpha} \ , \quad C_{base} = \sum_{n \in N} pb_n, \\
\theta &= \sum_{s \in S, i \in F_s} c_{f_s^i}, \lambda_{max} = \max_{n \in N} \frac{(pm_n - pb_n)}{c_n}, \\
\lambda_{min} &= \min_{n \in N} \frac{(pm_n - pb_n)}{c_n}
\end{aligned} \tag{26}
$$

$$
\frac{C}{C^*} \geq \frac{C_{base} \cdot \alpha}{C_{base} \cdot \alpha + \lambda_{max} \cdot \beta \cdot \sum\limits_{s \in S} \sum\limits_{i \in F_s} c_{f_s^i}} \tag{27}
$$

where $\left(C_{base} + \sum\limits_{s \in S, i \in F_s} c_{f_s^i} \cdot \lambda_{min}\right) \cdot \alpha$ represents the optimal cost for serving requested SFCs S. The cost is obtained, by placing the VNFs of the requested SFCs at a node with the lowest energy cost of computational resources per unit, and the node provides network services with green energy. The time complexity of the algorithm is analyzed as follows. For each VNF it takes $O(n^2)$ time to determine the candidate set and find out a node w with maximal bft_w, in line 7 of Algorithm 1, because it needs to determine the shortest distance and path from source o to other nodes by Dijkstra algorithm. Then, for a request $s \in S$, it consists of H_s VNFs. There are a total of $|S|$ requests to be placed and routed. Therefore, the time complexity of Algorithm 1 is $O\left(|S| \cdot \max\limits_{s \in S} H_s \cdot |N(t)|\right)$.

D. Availability-enhanced algorithm (AEA)

SPA places VNFs of SFCs decentrally, which increases the number of communication links of SFCs and decreases the availability of SFCs. So, for SFCs whose actual availability $\hat{R}_s$ does not meet expectations, AEA first uses the server where VNF of primal SFC is placed as the source point o, then select the server o' with the most remaining computational resources within the $|F_s|$ hop range of the source point o, and place VNFs of the SFC to this node as much as possible, which is shown in line 9. If there are m remaining VNFs that cannot be placed due to the limited capacity of server o', then within the m hop range of server o', find the next node with the largest remaining computational resources and feasible routing path for placing remaining VNFs of SFC and routing traffic, which is shown in lines 13-17. AEA repeats this process until all VNFs of the SFC are placed. During the placement of SFC, if there are limited computational resources that make it impossible to place VNF, or the p_{max} SFCs still cannot meet the availability requirement, the requested SFC s will be refused and be added to the next time slot, which is shown in lines 21-23. Otherwise, continue to place multiple SFCs until availability requirement are met.

For the requested SFC $s \in S$ that has satisfied the availability, AEA selects each VNF of each $p_i\,(1 \le p_i \le p_s(t))$ SFC sequentially and try to place the VNF to other node z. The green energy of node z is required to be greater than that of the previous node, and the availability of SFCs should still meet the expectation.

Algorithm 2: AEA

Input: $G(t) = (N(t), E(t))$, S
Output: $D1[i][j][h]$, $D2[i][j][h]$, $RS(t+1)$
1 **foreach** s *in* S **do**
2 $count = 1$;
3 **while** $\hat{R}_s < R_s$ **do**
4 $o \leftarrow D1[s][1][1]$, $count = count + 1$;
5 **if** $count > p_{max}$ **then**
6 $RS\,(t+1) \leftarrow RS\,(t+1) \cup s$;
7 Update resources for nodes and links;
8 **break;**
9 $o' \leftarrow$select node n with maximal $cap_n \ge c_{f_s^1}$ and $shortdistance(o, n) \le |F_s|$;
10 i = 1 ;
11 **while** $f_s^i \in F_s$ **do**
12 **if** $\exists o'$ **then**
13 **if** $cap_{o'} < c_{f_s^i}$ **then**
14 $o \leftarrow o'$;
15 $o' \leftarrow$ select node n with maximal $cap_n \ge c_{f_s^i}$ and $shortdistance(o, n) \le |F_s - i + 1|$;
16 Establish shortest path $\Psi_{o,o'}$ between o and o' with Dijkstra Algorithm, according to the traffic load of the SFC s and the communication resources of the network;
17 $D2[s][count][i-1] \leftarrow \Psi_{o,o'}$
18 **else**
19 $D1[s][count][i++] \leftarrow o'$;
20 update resources for nodes and links;
21 **else**
22 $RS\,(t+1) = RS\,(t+1) \cup s$;
23 Refuse s and Update resources for nodes and links;
24 **end**
25 **end**
26 Loop through each VNF and attempt to place it to nodes with higher green energy while SFC availability requirement is met.
27 **end**

E. AEA Algorithm analysis

The basic energy consumption of the node has been calculated in the first stage. Therefore, in the second stage, only the energy consumption cost caused by multiple SFCs placement needs to be calculated, and its optimal solution is 0, which means no additional backup SFCs deployment is required to enhance availability. Similarly, for a given request set S, the upper bound of energy costs for Algorithm 2 is $\beta \cdot p_{max} \cdot \lambda_{max} \cdot \sum_{s\in S} \sum_{i\in F_s} c_{f_s^i}$. To ensure service availability, AEA add $p_{max} - 1$ SFC at most for each request. For each VNF, it is necessary to spend at most $O(|N(t)|^2)$ to determine the placement node and shortest path, as it calls the Dijkstra algorithm on lines 14-16 of Algorithm 2. Furthermore, one SFC $s \in S$ has H_s VNFs, and $|S|$ SFCs need to be enhanced for availability. So, the time complexity of Algorithm 2 is $O\left(|S| \cdot p_{max} \cdot H_s \cdot |N(t)|^2\right)$.

F. OTGRSP scheme

At the beginning of each time slot t, OTGRSP updates network information based on vehicle data collected by the controller, which includes the number, locations, and battery status of nodes in the network, as well as the availability of links. Then, OTGRSP calculates the availability of SFCs and add SFCs that cannot meet the availability requirements to $RS(t)$, which is shown in lines 3-4. Next, the algorithm SPA takes network $G(t)$ and $RS(t)$ as inputs to solve the VNFs placement and traffic routing decisions, and returns SFCs $RS(t+1)$ that cannot be serviced, which is shown in lines 5. Due to the volatility of link availability and the uncertainty of green energy, the provided SFC may not be able to meet the availability requirements of the service and efficiently utilize green energy to reduce energy costs in the current time slot. So, Algorithm 2 takes $G(t)$ and $R\backslash RS(t+1)$ as inputs to improve the availability of SFCs and returns green reliable VNF placement and traffic routing decisions, as well as SFCs $RS(t+1)'$ that cannot be serviced. Finally, OTGRSP will attempt to provide services for $RS(t+1)$ in the next time slot and in the order of being added to $RS(t+1)$, which is shown in lines 7. Given the finite resources of edge nodes, it is inevitable that some user requests cannot be served if the network demand exceeds the capacity of the edge nodes.

Notes: (1) In larger GMENs with a higher number of edge nodes and VNFs, there is a way to easily scale OTGRSP to provide services for users. The way is to divide larger GMENs into multiple sub-networks, each sub-network includes some edge nodes and VNFs that constitute the SFC. Then, the OTGRSP is called in each sub-network to solve placement and routing decisions. The size of a sub-network can be adaptively adjusted based on the previous historical performance of OTGRSP. (2) In case of network congestion, OTGRSP rejects the underserved SFCs and process them in the next time slot. In the next time slot, SPA will resupply them based on the latest network environment. Since SPA is based on the latest network environment, it may take different decisions from the previous time slot. In other words, OTGRSP uses a "deny and reassign approach" to alleviate the problems caused by network congestion.

G. OTGRSP scheme performance analysis

Lemma 3. Given a green mobile edge network and a request set S within the time range T, the approximate scheme, OTGRSP, can provide an approximate solution for the LGRSP problem within $O\left(|T| \cdot |S| \cdot p_{max} \cdot H_s \cdot |N(t)|^2\right)$ time complexity. The upper and lower bound of approximate ratio of OTGRSP are $\frac{\beta\cdot(C_{base}+\eta_{max})}{\left(C_{base}+\lambda_{min}\cdot\sum_{s\in S,i\in F_s} c_{f_s^i}\right)\cdot\alpha}$ and

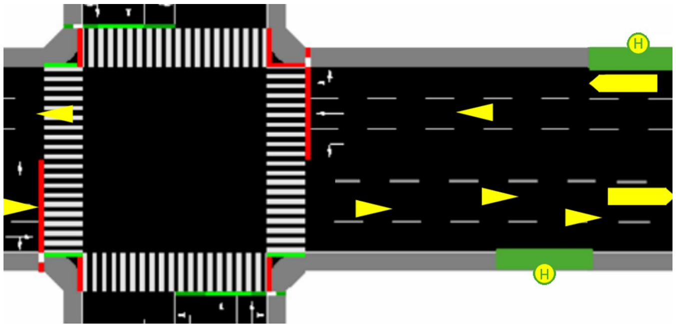


Fig. 4. Simulated Road

$\frac{\alpha \cdot C_{base}}{\beta \cdot (C_{base}+\eta_{max})}$, where $C_{base} = \sum\limits_{n \in N} pb_n, \eta_{max} = (1+p_{max}) \cdot \lambda_{max} \cdot \sum\limits_{s \in S} \sum\limits_{i \in F_s} c_{f_s^i}$, $\lambda_{min} = \min\limits_{n \in N} \frac{(pm_n - pb_n)}{c_n}$.

Algorithm 3: OTGRSP

Input: T, $G(t) = (N(t), E(t))$, S
Output: Green reliable SFC provisioning scheme

Initially, $RS(1) = S$;
foreach t *in* T **do**
- Update network topology information, which includes the number, location, and battery state of nodes, and the availability of link at t time slot;
- Add requests that cannot meet availability requirements at t time slot to $RS(t)$;
- VNF placement and traffic routing decisions for $RS(t)$ and $RS(t+1)$ are obtained by Algorithm1 with input $G(t)$, $RS(t)$;
- Reliable VNF placement and traffic routing decisions for $R \setminus RS(t+1)$ and $RS'(t+1)$ are obtained by Algorithm2 with input $G(t)$, $R \setminus RS(t+1)$;
- $RS(t+1) = RS(t+1) \cup RS'(t+1)$;

end

Proof. At the beginning of each time slot $t \in T$, it needs to spend $O\left(|N(t)|^2\right)$ time updating network information because it needs to calculate the availability of the link between two nodes. Next, it needs to spend $O(p_{max} \cdot H_s \cdot |N(t)|)$ to calculate the actual availability of each request, because calculating the availability of each service function chain is $O\left(H_s + (|N(t)|-1) \cdot (H_s - 1)\right)$. Where, $O\left(H_s\right)$ represents the time to calculate the availability of H_s VNFs. $O\left(|N(t)|-1\right)$ represents the time to calculate the availability of links between a VNF pair, that is, the shortest path length with the maximum hop count between two nodes in a green mobile edge network. In addition, the time complexity of Algorithm 1 and Algorithm 2 is $O\left(|S| \cdot H_s \cdot |N(t)|^2\right)$ and $O\left(|S| \cdot p_{max} \cdot H_s \cdot |N(t)|^2\right)$, respectively. Therefore, the time complexity of the OTGRSP algorithm is $O\left(|T| \cdot |S| \cdot p_{max} \cdot \max\limits_{s \in S} H_s \cdot |N(t)|^2\right)$. Next, we will prove the approximation ratio.

$$
\begin{aligned}
\frac{C_{all}}{C_{all}^*} &\leq \frac{|T| \cdot \left(C' + C''\right)}{C_{all}^*} \leq \frac{|T| \cdot \left(C' + C''\right)}{|T| \cdot \left(\sum\limits_{n \in N} pb_n + \lambda_{min} \cdot \sum\limits_{s \in S, i \in F_s} c_{f_s^i}\right) \cdot \alpha} \\
&\leq \frac{\beta \cdot \left(C_{base} + \lambda_{max} \cdot \sum\limits_{s \in S, i \in F_s} c_{f_s^i}\right) + \beta \cdot \left(p_{max} \cdot \lambda_{max} \cdot \sum\limits_{s \in S, i \in F_s} c_{f_s^i}\right)}{\left(C_{base} + \lambda_{min} \cdot \sum\limits_{s \in S, i \in F_s} c_{f_s^i}\right) \cdot \alpha} \\
&\leq \frac{\beta \cdot (C_{base} + \eta_{max})}{\left(C_{base} + \lambda_{min} \cdot \sum\limits_{s \in S, i \in F_s} c_{f_s^i}\right) \cdot \alpha}, \\
\lambda_{min} &= \min_{n \in N} \frac{(pm_n - pb_n)}{c_n}, C_{base} = \sum_{n \in N} pb_n, \\
\eta_{max} &= (1+p_{max}) \cdot \lambda_{max} \cdot \sum_{s \in S} \sum_{i \in F_s} c_{f_s^i}
\end{aligned}
\tag{28}
$$

where C_{all} is the cost of OTGRSP strategy. C_{all}^* is the optimal solution to the problem. C' and C'' respectively represent the upper bound of the energy cost spent by the first and second stage algorithms. They are not related to time. So, the performance guarantee of each time slot can deduce the theoretical bound of the algorithm for multiple time slot problems. On the other hand, in the presence of feasible solutions, the lower bound on the approximation ratio of OTGRSP is that,

$$
\begin{aligned}
\frac{C_{all}}{C_{all}^*} \geq & \frac{\sum\limits_{n \in N} pb_n \cdot \alpha \cdot |T|}{C_{all}^*} \\
& \geq \frac{\alpha \cdot C_{base}}{\beta \cdot \left(C_{base} + p_{max} \cdot \lambda_{max} \cdot \sum\limits_{s \in S, i \in F_s} c_{f_s^i}\right)} \\
& \geq \frac{\alpha \cdot C_{base}}{\beta \cdot (C_{base} + \eta_{max})}
\end{aligned}
\tag{29}
$$

where $\beta \cdot \left(C_{base} + p_{max} \cdot \lambda_{max} \cdot \sum\limits_{s \in S, i \in F_s} c_{f_s^i}\right)$ represents the largest energy costs for serving S requested SFCs. The cost is obtained based on two operations. The first operation is to place the VNFs of the requested SFCs at a node with the highest energy cost of computational resources per unit. The node provides network services with backup power supply. The second operation is to provide p_{max} SFCs for each requested SFC to guarantee service availability.

V. PERFORMANCE EVALUATION

In this section, we first describe the experimental setting of the problem and the benchmark schemes. Then, a green mobile edge network is constructed based on a simulated dataset. Finally, the performance of different scenarios is evaluated and the impact of important parameters is analyzed. All experiments are run on a personal computer equipped with, a Python-based simulator, 3.20GHz Intel(R) Core(TM)I5-11320H CPU and 16GB RAM.

A. Environment setting

In this paper, we use SUMO and MOVE simulation tools [37] to generate moving trajectories of vehicles in a realistic urban environment and select vehicles on a two-way urban road covered by a base station as mobile edge nodes, as shown in Fig.4. The setup of building a green mobile edge network with a single base station area is easily scalable to the study of multiple areas. The initial position of the vehicle is randomly located on the city map, and it will move along the street in a fixed direction. The characteristics of mobile edge nodes are as follows: (1) The number of CPU cores is 16 [8]; (2) The communication range is 200 meters [11]; (3) The availability is set to be evenly distributed within the range of [0.999,0.9999]. The idle and full load energy consumption of the nodes is 170W and 500W respectively [20], the capacity of the battery is set to 5KWh, and the bandwidth of the inter-node communication link is 10Mbps. Due to the lack of datasets of energy harvested by solar cars and the fact that the variation curve of solar radiation intensity during the day is similar to a normal distribution, i.e., the intensity is weak in the morning and evening and high in the middle of the day. We assume that the green energy harvested by the mobile edge nodes obeys a normal distribution with expectation of 250W which is half of the full load energy consumption. The cost factor of the green energy is $\alpha = 0.5$, and the cost factor of the backup power supply (e.g., diesel generation) is $\beta = 10\alpha$ [35]. A total of 5 different virtual network functions have been set up in the network, and their required computational resources are 2, 1, 1, 2, and 3 CPU cores, respectively. We randomly generate SFCs requested by users, with availability requirement evenly distributed in the range of $[0.9, 0.999]$, and bandwidth requirement evenly distributed in the range of $[2Mbps, 5Mbps]$.

B. Benchmark

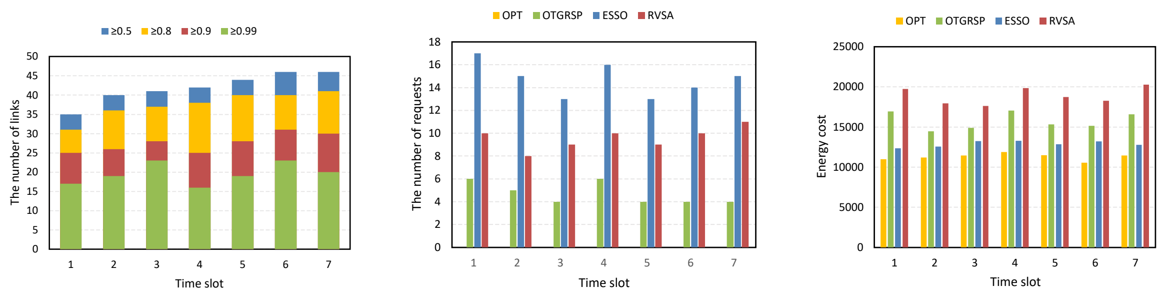
Before presenting the performance evaluation, we briefly describe the compared algorithms. OTGRSP scheme: In order to solve the LGRSP problem, this paper proposes the OTGRSP scheme. Optimal scheme (OPT): It solves LGRSP problem by means of a Lingo solver, which is able to provide an optimal solution to the problem. Energy Smart Service Function Chain Orchestrator (ESSO) [8]: It is heuristic algorithm which selects a path with the most green energy to route the data from a candidate set containing k shortest paths and uses dynamic programming and taboo-based search algorithms to solve VNF deployment decisions in this path to minimize the cost. Recursive VNF Scheduling Algorithm (RVSA) [12]: It is a recursive based heuristic algorithm which iteratively places the VNFs for each SFC. In case a node is unable to place the VNF due to insufficient processing capacity, it returns to the previous iteration and places the VNF again. The data routing path between the nodes is the shortest path that satisfies the requested bandwidth requirement. The program does not end until all VNFs are placed or exceed the given runtime.

C. Performance evaluation

1) The impact of network topology changes on the performance of different schemes. This experiment is based on the speed and location of mobile edge nodes in each time slot in $T = [1, 7]$ to update the green mobile edge network. In this time interval, the number of mobile edge nodes is 16 and the number of requested SFCs is 30 and the length of each SFC is 3. The number of links whose availability exceeds 0.5, 0.8, 0.9, and 0.99 in seven time slots for the network topology is shown in Fig.5a. It can be observed that the number of links exceeding the given threshold in different time slots is variable and unstable. The reason is that in reality the quality of communication links between mobile edge nodes is easily affected by buildings and distance between nodes etc.

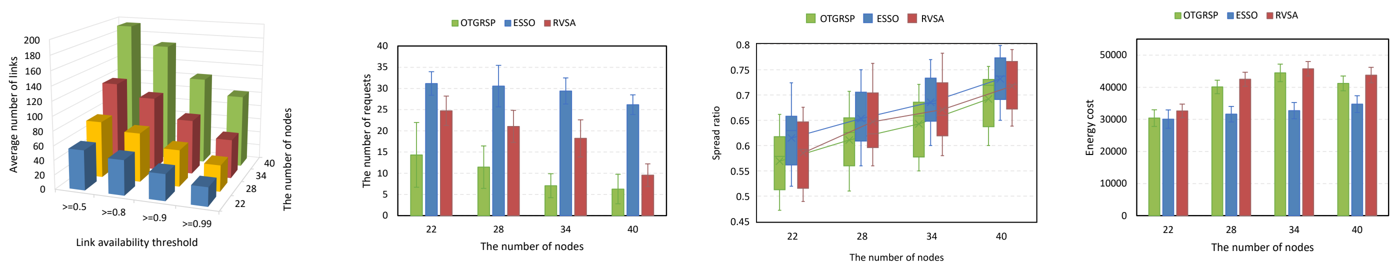
Fig. 5b shows the number of SFCs violating the availability requirement for the four strategies in each time slot, and the average number of availability violations for the OPT, OTGRSP, ESSO, and RVSA strategies are 0, 4.7, 14.7, and 9.5, respectively. The ESSO strategy does not consider the availability of nodes and links as well as provide a protection scheme, which makes it susceptible to violating the user's availability requirements. The RVSA scheme simply selects the shortest link between two nodes to construct the routing path, and it ignores the availability of the link which is lower and unstable compared to the availability of nodes in GMENs. The OTGRSP compensates for the shortcomings of the ESSO and RVSA schemes by making a trade-off between the green energy harvested by the nodes and the availability of the links. So, OTGRS violates the availability requirement with a lower number of requests. Because OTGRSP forces VNF to be deployed on different nodes in the first stage algorithm, making the primal SFC more susceptible to link interference, which is why it is inferior to the optimal OPT algorithm.

Fig. 5c reacts the energy cost of the four strategies, OPT, OTGRSP, ESSO and RVSA, at each time slot, and their average values are 11285.71, 15774.14, 12892.85, and 18919.21, respectively. The average energy cost of the OTGRSP strategy is 0.8 times higher than that of the RVSA. The reason is that the RVSA scheme does not consider the disparity of green energy harvested by different nodes, and it also needs to provide additional backup SFCs to improve the availability of the network. The average energy cost of the OTGRSP strategy is 1.22 times higher than that of ESSO. Although the average energy cost of the OTGRSP strategy is 22% higher than that of ESSO, the average number of requests for which it meets the service availability requirements is 1.6 times higher than that of ESSO. The reason is that the OTGRSP scheme requires additional deployment of backup SFCs to improve availability. Because OTGRSP adopts a fixed strategy in the first and second stages, that is, forcibly placing the VNF of SFC on different nodes and as much as possible placing the VNF on one node, it cannot make the optimal trade-off, which is the reason why it is inferior to the optimal OPT algorithm. Since this problem is an integer nonlinear programming problem with concatenated multiplication, the Lingo solver is slow and may not be a globally optimal solution under large scale conditions. In the following we only compare the performance



(a) Number of links exceeding a given availability threshold (b) Number of SFCs violating the availability requirement (c) Energy cost

Fig. 5. The impact of network topology changes on the performance of different schemes.



(a) Average number of links exceeding given availability thresholds (b) Average number of SFCs violating the availability requirement (c) Average spread rate of SFCs (d) Average energy costs

Fig. 6. Impact of the number of mobile edge nodes on the performance of different schemes.

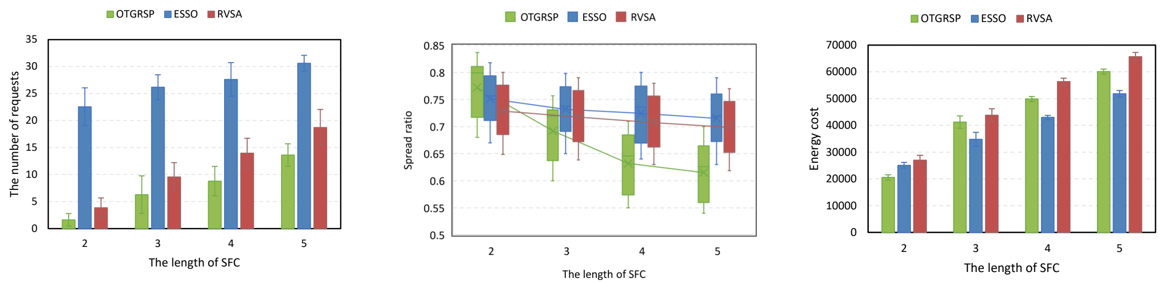
of the other three strategies.

2) Impact of the number of mobile edge nodes on the performance of different schemes. Fig. 6 shows the performance results of the three strategies for the case of increasing the number of mobile edge nodes from 22 to 40, which is obtained from an experiment of supplying 50 SFCs of length 3 in 20 time slots. Fig. 6a reflects the relationship between the number of nodes and the average number of links exceeding a given link availability threshold in 20 time slots. From the chart, it can be seen that the average number of links exceeding the given availability threshold increases with the increase in the number of nodes. The reason is that as the number of nodes increases, road traffic becomes denser, i.e. the average distance between vehicles shortens, so the availability of communication links is higher.

Fig. 6b is a bar chart, which reflects that as the number of nodes increases, i.e., the denser the road is, there is a decrease in the average number of SFCs violating the availability requirement in the time interval. The reason is that on denser roads, the availability of communication links is higher, as shown in Fig. 6a. Moreover, as the number of mobile edge nodes increases, the network's computational and communication resources are more sufficient. So, OTGRSP and RVSA can provide more backup SFCs to enhance the availability of SFCs. However, OTGRSP outperforms RVSA and ESSO. The reason is as follows. ESSO doesn't consider and enhance the availability of SFC. On the other hand, RVSA is less efficient than OTGRSP in improving the availability of SFCs. RVSA provides only P_{max} backup SFCs for SFCs by a policy. The policy doesn't consider the difference in the availability of links and simply uses the shortest path between nodes as the route. And, VNFs of backup SFCs are placed on some nodes in a decentralized manner. Instead, AEA tries to avoid placing the VNFs of the backup SFCs at different nodes, thus reducing the impact of link failures on the availability of the backup SFCs.

Fig. 6c reflects the number of nodes and the average spread rate of SFC in 20 time slots. Here, the spread rate of SFC is the ratio of the number of servers occupied by an SFC to the number of VNFs present in that SFC. This measures the distribution of SFCs over servers and evaluates the green energy exploration and load balancing capabilities of different strategies [8]. As the number of nodes increases, all the three strategies prefer to place the VNFs of the SFC to different nodes. The reason is that, with a fixed number of VNFs in the SFC, the higher the number of nodes, the lower the probability for the RVSA scheme to deploy VNFs to previously deployed nodes. For the ESSO scheme, the higher the number of network nodes and the higher the number of nodes in its path to find the maximum green energy, the higher the chances of placing VNFs to different nodes in a decentralized manner. For the OTGRSP scheme, the denser the number of nodes, the higher availability of the links, which makes it more inclined to decentralize the placement of VNFs. However, in order to improve the availability of the SFC, the AEA algorithm places the backup SFC on a single node as much as possible, which makes its spread rate relatively lower than the other two strategies.

Fig. 6d reflects the number of nodes and the average energy cost in 20 time slots. The energy cost of the three strategies increases with the number of nodes, but after 34 nodes the energy cost of the OTGRSP and RVSA strategies decreases. This is due to the fact that each mobile edge node has its own base energy consumption and the capacity of the green

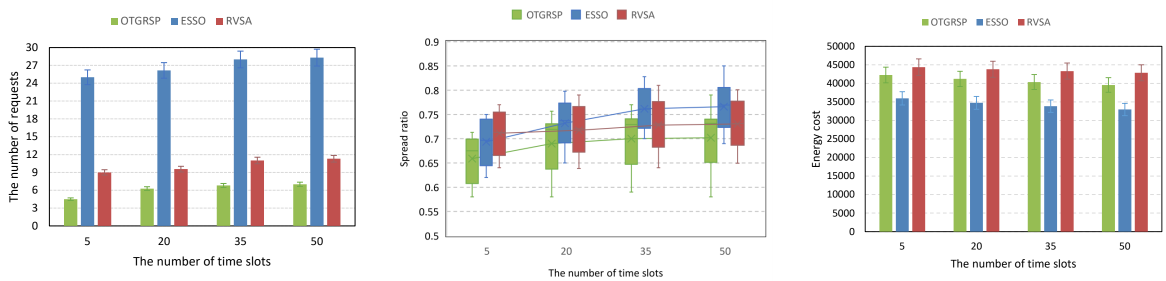


(a) Average number of SFCs violating the availability requirement

(b) Average spread rate of SFCs

(c) Average energy costs

Fig. 7. Impact of SFC length on the performance of different schemes.



(a) Average number of SFCs violating the availability requirement

(b) Average spread rate of SFCs

(c) Average energy costs

Fig. 8. Impact of the number of time slots on the performance of different schemes.

energy utilized by the three strategies is not enough to offset the introduced energy consumption, so as the number of nodes increases thus making the energy overhead of the network larger. Secondly, as the number of nodes increases, the availability of network links improves, and the number of backup SFCs that need to be provided by the OTGRSP and RVSA decreases, which results in some relative reduction in energy costs. The energy cost of the OTGRSP and RVSA schemes is higher than that of the ESSO scheme because both of them need to provide backup SFCs. The RVSA scheme does not provide SFC based on the green energy harvested by nodes, which results in its energy consumption cost being higher than that of OTGRSP and ESSO. In the case of 22 nodes, the energy cost of OTGRSP and ESSO schemes is almost equal because the nodes have limited computational resources and OTGRSP can only backup a few SFCs. Secondly, the OTGRSP scheme has green energy storage capability and its ability to utilize green energy is slightly greater than the ESSO scheme.

3) Impact of SFC length on the performance of different schemes. Fig. 7 shows the performance results of the three strategies for increasing the length of the SFC from 2 to 5, which is obtained from an experiment of supplying 50 SFC requests in 40 edge nodes within 20 time slots. Fig. 7a reflects the number of requests violating the availability requirements tends to increase as the length of the SFC increases. This is because the unavailability of any node or link in the SFC leads to the unavailability of the SFC. OTGRSP takes link unavailability into account and sets a penalty factor when placing VNFs, so that it have fewer SFCs violating availability requirements than the RVSA, given the maximum number of backups.

Fig. 7b reflects the length of SFCs and the average spread rate of SFC in 20 time slots. The spread rate of the three strategies decreases as the length of the SFC increases. The reason is that with a fixed number of nodes, placing more VNFs clearly increases the chances of placing VNFs to the same nodes. As the SFC length is longer, the SFC is more prone to failure, which makes the AEA algorithm of OTGRSP scheme need to provide more backup SFCs and are more inclined to be placed on the same node to reduce the impact of link unavailability. Therefore, the spread rate of the OTGRSP scheme is lower than the other two schemes when the SFC length is greater than or equal to 3.

Fig. 7c reflects the length of SFC and the average energy cost in 20 time slots. The energy cost of the three strategies increases as the SFC length increases. This is because as the SFC length increases, the number of VNFs that need to be placed also increases, which increases the network energy requirement. The energy cost of the OTGRSP algorithm is smaller than ESSO when the SFC length is less than 3. The reason is that, OTGRSP has a higher spread rate than the other two and can more fully utilize the green energy harvested by nodes. At SFC length greater than or equal to 3, the OTGRSP and RVSA schemes need to additionally provide backup SFCs, which takes up computational resources and consumes energy, which results in a greater energy cost for both than the ESSO strategy.

4) Impact of the number of time slots on the performance of different schemes. Fig. 8 shows the performances of the three strategies, for four different time slots ranging from 5 to 50. These results were obtained by performing an experiment where 50 SFC requests were supplied in 40 edge nodes within 20 time slots. Fig. 8a reflects an increasing trend in the number of SFCs violating the availability requirement as the number of time slots increases. The reason for this is that the location of nodes is more likely to change significantly over longer time intervals, which will negatively affect the availability of the wireless communication links. Fig. 8b and Fig. 8c show an increasing trend in the spread rate and a decreasing trend in the energy cost, as the number of time slots increases. The reason for this is that in a time slot, the green energy harvested by some nodes may not be fully utilized and the green energy will be stored in batteries. In the subsequent time slots, the three strategies are more likely to place VNFs in a decentralized manner to further utilize the green energy, reducing the energy cost of network services. Since RVSA does not perceive green energy to place the VNFs, its performance changes in terms of diffusion rate and energy cost are insignificant.

Overall, the results of experiments indicate that OTGRSP ensures service availability while reducing the energy cost of all edge nodes and has better performance than other algorithms. The reason is that OTGRSP is specialized for solving the LGRSP problem. The problem is first presented in our paper. Other algorithms are not suitable for solving the problem because they do not comprehensively consider the challenges associated with this problem.

VI. Conclusion

In this paper, we first propose the LGRSP problem, which considers the dynamic nature of green mobile edge networks, such as the number, location, and link availability of edge nodes, as well as green energy. Then, we discretize the problem into a series of one-slot problem OGRSP and prove that both types of problems are NP-hard. Next, we propose an approximate algorithm OTGRSP, which can make a trade-off between ensuring SFC availability and utilizing green energy. Finally, we conducted extensive experiments using the dataset generated by the simulator to evaluate the performance of the proposed algorithm. The results indicate that OTGRSP has better performance than other state-of-the-art methods.

References

[1] P. Zhou, C. Wang, and Y. Yang, "Design and optimization of solar-powered shared electric autonomous vehicle system for smart cities," *IEEE Transactions on Mobile Computing*, 2021.

[2] Innovative solar technology for transport logistics, 2023. [Online]. Available: https://sonomotors.com/en/solar-integration/

[3] Y. Sun, D. Xu, D. W. K. Ng, L. Dai, and R. Schober, "Optimal 3d-trajectory design and resource allocation for solar-powered uav communication systems," *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 4281–4298, 2019.

[4] S. Sekander, H. Tabassum, and E. Hossain, "Statistical performance modeling of solar and wind-powered uav communications," *IEEE Transactions on Mobile Computing*, vol. 20, no. 8, pp. 2686–2700, 2021.

[5] V. R. Chintapalli, B. R. Killi, R. Partani, B. R. Tamma, and C. S. R. Murthy, "Energy- and reliability-aware provisioning of parallelized service function chains with delay guarantees," *IEEE Transactions on Green Communications and Networking*, pp. 1–1, 2023.

[6] J. Liang and F. Tian, "An online algorithm for virtualized network function placement in mobile edge industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 3, pp. 2496–2507, 2022.

[7] Y. Qiu, J. Liang, V. C. Leung, X. Wu, and X. Deng, "Online reliability-enhanced virtual network services provisioning in fault-prone mobile edge cloud," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 7299–7313, 2022.

[8] M. F. Bari, S. R. Chowdhury, and R. Boutaba, "Esso: An energy smart service function chain orchestrator," *IEEE Transactions on Network and Service Management*, vol. 16, no. 4, pp. 1345–1359, 2019.

[9] B. Mao, F. Tang, Y. Kawamoto, and N. Kato, "Ai models for green communications towards 6g," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 210–247, 2021.

[10] J. Bai, X. Chang, F. Machida, L. Jiang, Z. Han, and K. S. Trivedi, "Impact of service function aging on the dependability for mec service function chain," *IEEE Transactions on Dependable and Secure Computing*, 2022.

[11] X. Zhang, X. Cao, L. Yan, and D. K. Sung, "A street-centric opportunistic routing protocol based on link correlation for urban vanets," *IEEE Transactions on Mobile Computing*, vol. 15, no. 7, pp. 1586–1599, 2016.

[12] S. Yang, F. Li, R. Yahyapour, and X. Fu, "Delay-sensitive and availability-aware virtual network function scheduling for nfv," *IEEE Transactions on Services Computing*, vol. 15, no. 1, pp. 188–201, 2022.

[13] J. Xu, L. Chen, and S. Ren, "Online learning for offloading and autoscaling in energy harvesting mobile edge computing," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 3, pp. 361–373, 2017.

[14] W. Chen, D. Wang, and K. Li, "Multi-user multi-task computation offloading in green mobile edge cloud computing," *IEEE Transactions on Services Computing*, vol. 12, no. 5, pp. 726–738, 2019.

[15] Y.-J. Ku, P.-H. Chiang, and S. Dey, "Real-time qos optimization for vehicular edge computing with off-grid roadside units," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 11 975–11 991, 2020.

[16] Y.-J. Ku, S. Baidya, and S. Dey, "Adaptive computation partitioning and offloading in real-time sustainable vehicular edge computing," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 12, pp. 13 221–13 237, 2021.

[17] Y. Qiu, J. Liang, V. C. Leung, and M. Chen, "Online security-aware and reliability-guaranteed ai service chains provisioning in edge intelligence cloud," *IEEE Transactions on Mobile Computing*, pp. 1–16, 2023.

[18] A. Hmaity, M. Savi, F. Musumeci, M. Tornatore, and A. Pattavina, "Virtual network function placement for resilient service chain provisioning," in *2016 8th International Workshop on Resilient Networks Design and Modeling (RNDM)*. IEEE, 2016, pp. 245–252.

[19] L. Qu, C. Assi, K. Shaban, and M. J. Khabbaz, "A reliability-aware network service chain provisioning with delay guarantees in nfv-enabled enterprise datacenter networks," *IEEE Transactions on Network and Service Management*, vol. 14, no. 3, pp. 554–568, 2017.

[20] L. Mai, Y. Ding, X. Zhang, L. Fan, S. Yu, and Z. Xu, "Energy efficiency with service availability guarantee for network function virtualization," *Future Generation Computer Systems*, vol. 119, pp. 140–153, 2021.

[21] X. Shang, Y. Liu, Y. Mao, Z. Liu, and Y. Yang, "Greening reliability of virtual network functions via online optimization," in *2020 IEEE/ACM 28th International Symposium on Quality of Service (IWQoS)*. IEEE, 2020, pp. 1–10.

[22] X. Shang, Y. Huang, Z. Liu, and Y. Yang, "Reducing the service function chain backup cost over the edge and cloud by a self-adapting scheme," *IEEE Transactions on Mobile Computing*, vol. 21, no. 8, pp. 2994–3008, 2021.

[23] B. Németh, N. Molner, J. Martín-Pérez, C. J. Bernardos, A. de la Oliva, and B. Sonkoly, "Delay and reliability-constrained vnf placement on mobile and volatile 5g infrastructure," *IEEE Transactions on Mobile Computing*, vol. 21, no. 9, pp. 3150–3162, 2022.

[24] J. Martín-Pérez, F. Malandrino, C. F. Chiasserini, M. Groshev, and C. J. Bernardos, "Kpi guarantees in network slicing," *IEEE/ACM Transactions on Networking*, vol. 30, no. 2, pp. 655–668, 2022.

[25] S. G. Kulkarni, M. Arumaithurai, K. Ramakrishnan, and X. Fu, "Rearm: Renewable energy based resilient deployment of virtual network functions," in *2017 European Conference on Networks and Communications (EuCNC)*. IEEE, 2017, pp. 1–6.

[26] J. Wang, B. He, J. Wang, and T. Li, "Intelligent vnfs selection based on traffic identification in vehicular cloud networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4140–4147, 2018.

[27] W. Zhuang, Q. Ye, F. Lyu, N. Cheng, and J. Ren, "Sdn/nfv-empowered future iov with enhanced communication, computing, and caching," *Proceedings of the IEEE*, vol. 108, no. 2, pp. 274–291, 2020.

[28] J. Bai, Y. Li, X. Chang, F. Machida, and K. S. Trivedi, "Understanding nfv-enabled vehicle platooning application: A dependability view," *IEEE Transactions on Cloud Computing*, 2023.

[29] M. Pourghasemian, M. R. Abedi, S. S. Hosseini, N. Mokari, M. R. Javan, and E. A. Jorswieck, "Ai-based mobility-aware energy efficient resource allocation and trajectory design for nfv enabled aerial networks," *IEEE Transactions on Green Communications and Networking*, vol. 7, no. 1, pp. 281–297, 2023.

[30] J. Bai, X. Chang, R. J. Rodríguez, K. S. Trivedi, and S. Li, "Towards uav-based mec service chain resilience evaluation: A quantitative modeling approach," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 4, pp. 5181–5194, 2023.

[31] M. H. Eiza and Q. Ni, "An evolving graph-based reliable routing scheme for vanets," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 4, pp. 1493–1504, 2013.

[32] M. Hashem Eiza, T. Owens, Q. Ni, and Q. Shi, "Situation-aware qos routing algorithm for vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 12, pp. 5520–5535, 2015.

[33] P. Zhou, C. Wang, and Y. Yang, "Design and optimization of electric autonomous vehicles with renewable energy source for smart cities," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 1399–1408.

[34] M. S. Bute, P. Fan, L. Zhang, and F. Abbas, "An efficient distributed task offloading scheme for vehicular edge computing networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 12, pp. 13 149–13 161, 2021.

[35] H. Ma, P. Huang, Z. Zhou, X. Zhang, and X. Chen, "Greenedge: Joint green energy scheduling and dynamic task offloading in multi-tier edge computing systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 4, pp. 4322–4335, 2022.

[36] W. Liang, Y. Ma, W. Xu, Z. Xu, X. Jia, and W. Zhou, "Request reliability augmentation with service function chain requirements in mobile edge computing," *IEEE Transactions on Mobile Computing*, vol. 21, no. 12, pp. 4541–4554, 2022.

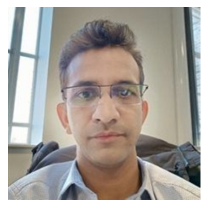
[37] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wiessner, "Microscopic traffic simulation using sumo," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2575–2582.

Junbin Liang received B.E. and M.S. degrees from Guangxi University, Nanning, China, and the Ph.D. degree from Central South University, Changsha, China, in 2000, 2005, and 2010, respectively. He was a visiting professor in University of British Columbia from 2019 to 2020. He is currently a professor at Guangxi University, Nanning, China. His research interests include sensor-cloud systems, fog computing, and distributed computing.

Shaodong Huang has been studying for the M.S. degree in electronic and information engineering at Guangxi University, Nanning, China, since 2021. His research directions include mobile edge clouds, network function virtualization, and optimization theory.

Yu Qiu received the BSc degree from the Tianjin University of Technology, Tianjin, China, in 2020, the ME degree in computer technology at the Guangxi University, Nanning, China, in 2023. He is currently working toward the Ph.D. degree in artificial intelligence at the South China University of Technology, Guangzhou, China. His research interests include metaverse, edge intelligence, network function virtualization, and optimization theory.

Lu Liu received the MSc degree in data communication systems from Brunel University, United Kingdom, and the PhD degree from the University of Surrey, United Kingdom. He is Professor of Data Science at the University of Leicester, United Kingdom. His research interests are in the areas of AI, cloud computing, service computing, and the Internet of Things, He has over 250 scientific publications in reputable journals and international conferences and as secured over 30 grants which are supported by UKRI/EPSRC, Innovate UK, Royal Society, British Council and leading industries. He is a Fellow of British Computer Society (BCS).



Furqan Aziz a lecturer of computer science at the University of Leicester. He received a PhD in computer Science from University of York, UK. His current research interest includes graph theory, network-based machine learning, computer vision, and bioinformatics.

Min Chen (Fellow, IEEE) has been a Full Professor with School of Computer Science and Engineering, South China University of Technology. He is also the director of Embedded and Pervasive Computing (EPIC) Lab at Huazhong University of Science and Technology. He is the founding Chair of IEEE Computer Society Special Technical Communities on Big Data. He was an assistant professor in School of Computer Science and Engineering at Seoul National University before he joined HUST. He is the Chair of IEEE Globecom 2022 eHealth Symposium. His Google Scholar Citations reached 40,000+ with an h-index of 96. His top paper was cited 4,304+ times. He was selected as Highly Cited Researcher from 2018 to 2022. He got IEEE Communications Society Fred W. Ellersick Prize in 2017, the IEEE Jack Neubauer Memorial Award in 2019, and IEEE ComSoc APB Oustanding Paper Award in 2022. He is a Fellow of IEEE and IET.