Journal of Systems Engineering and Electronics
Vol. 35, No. 4, August 2024, pp.829 – 841

Multi-network-region traffic cooperative scheduling in large-scale LEO satellite networks

LI Chengxi[1], WANG Fu[1,*], YAN Wei[2], CUI Yansong[1], FAN Xiaodong[3], ZHU Guangyu[2], XIE Yanxi[4], YANG Lixin[3], ZHOU Luming[1], ZHAO Ran[1], and WANG Ning[1]

1. School of Electronic Engineering, Beijing University of Posts and Telecommunications , Beijing 100876, China;
2. The 6th Research Institute of China Electronics Corporation, Beijing 102209, China; 3. China Satellite Network Share Operation Co., Ltd, Beijing 100876, China; 4. Weihai Electronic Information Technology Comprehensive Research Center of the Ministry of Industry and Information Technology Weihai 264200, China

Abstract: A low-Earth-orbit (LEO) satellite network can provide full-coverage access services worldwide and is an essential candidate for future 6G networking. However, the large variability of the geographic distribution of the Earth's population leads to an uneven service volume distribution of access service. Moreover, the limitations on the resources of satellites are far from being able to serve the traffic in hotspot areas. To enhance the forwarding capability of satellite networks, we first assess how hotspot areas under different load cases and spatial scales significantly affect the network throughput of an LEO satellite network overall. Then, we propose a multi-region cooperative traffic scheduling algorithm. The algorithm migrates low-grade traffic from hotspot areas to coldspot areas for forwarding, significantly increasing the overall throughput of the satellite network while sacrificing some latency of end-to-end forwarding. This algorithm can utilize all the global satellite resources and improve the utilization of network resources. We model the cooperative multi-region scheduling of large-scale LEO satellites. Based on the model, we build a system testbed using OMNET++ to compare the proposed method with existing techniques. The simulations show that our proposed method can reduce the packet loss probability by 30% and improve the resource utilization ratio by 3.69%.

Keywords: low-Earth-orbit (LEO) satellite network, satellite communication, load balance, multi- region scheduling, latency optimization.

DOI: 10.23919/JSEE.2024.000045

1. Introduction

With the increasing demand for communication capabilities from the Internet of Things, autonomous driving, and ocean shipping, low-Earth-orbit (LEO) satellite networks have become an essential complement to terrestrial networks, providing ubiquitous access to users worldwide [1,2]. Projects, such as, the Starlink program, the Iridium system, and the Globalstar cellular system, have established a large-scale LEO satellite network with thousands of satellites covering the globe [3]. However, due to the characteristics of the distribution of the Earth's land and oceans, most of the satellite-terrestrial services are concentrated in hotspot cities on the land [4]. Meanwhile, the ocean, which accounts for 70% of Earth's area, is covered with many LEO satellites but lacks satellite access services. As a result, most LEO satellites are currently under-used, while others suffer from a large amount of service traffic. The severely uneven distribution of services leads to a high cost and limited carrying capacity of satellite networks, compared to terrestrial networks, limiting the development of large-scale LEO satellite networks.

LEO satellite networks are located in orbits at altitudes between 160 km and 2 000 km, unlike terrestrial network nodes with a large bandwidth [5]. The inter-satellite links (ISLs) and bandwidth resources of existing satellite nodes with limited energy consumption, volume, and power, make it challenging to provide sufficient forwarding capacity for hotspot area services [6,7]. Satellite networks face enormous challenges relative to large cities with millions of people. Much of the existing research on LEO satellite networks is devoted to mobility management and congestion avoidance to improve service reliability. Regarding mobility management, existing studies use snapshot time domain graphs to reduce the drawbacks associated with the high-speed movement of satellites. Meanwhile, existing satellite networks enhance network robustness through agile handovers between the

---

Manuscript received March 27, 2023.
*Corresponding author.
This work was supported by the National Key R&D Program of China (2021YFB2900604).

satellites and the terrestrial links (STLs) [8]. As a result, satellite network's routing convergence speed and data forwarding efficiency have been effectively improved. However, it is difficult to effectively enhance the constrained resources of satellite networks, including computation, storage, and bandwidth. Furthermore, these cause inevitable network congestion in hotspot areas [9]. Besides, much research has been dedicated to network congestion avoidance, which has two main aspects. On the one hand, researchers avoid the aggravation of network congestion by detecting congestion areas and then de-congesting them for later traffic. On the other hand, congestion areas can be avoided by balancing the network load through flexible routing algorithms. Both approaches provide beneficial effects in large-scale LEO satellite networks [10,11].

Among the approaches to avoid network congestion, congestion-aware methods can only provide countermeasures after congestion occurs and can hardly alleviate congestion in hotspots. Therefore, avoiding network congestion with a suitable load-balancing routing approach has become a proven method. For load balancing routing, back-pressure based routing is a widely studied approach for LEO satellite networks to tackle the resource allocation problem and prevent traffic congestion [12,13]. However, the load balancing method based on backpressure routing finds it difficult to counter the pressure in real-time in multi-ISL and large-scale space scenarios. Therefore, it is difficult to deploy these methods in large-scale LEO networks. In addition, multi-path collaborative routing is considered an attractive approach to improve transmission reliability [14]. However, while improving the utilization of free resources, the large number of redundant packets imposes an additional burden on an LEO satellite network. Providing more network states can improve the accuracy of the routing algorithms, as has been demonstrated in many studies. However, obtaining network resources in real-time at large spatial scales is challenging when the onboard resources are as constrained as they are [15]. In addition, many studies use cooperative scheduling of LEOs with medium-Earth orbits (MEOs), LEOs and ground gateway stations to achieve network load balancing. However, it is difficult for these studies to provide a uniform model to account for the additional overhead that introducing heterogeneous nodes would impose on the LEO satellite network [16,17]. Therefore, utilizing large-scale LEO satellites to avoid network congestion and improve throughput still needs to be studied.

To address these problems, we consider grooming overflowed traffic into cold areas for transmission, which balances the load in the routing computing stage. For the traffic hotspot areas, the proposed algorithm reasonably diverts the traffic to improve the overall service capability of the network. While solving the load imbalance problem [15] of the LEO satellite network, our method can discriminate between traffic flows with different qualities of service (QoS) and perform multi-level transmission approaches. Besides, the method can dynamically control the service forwarding range according to the service priority. Therefore, this method can improve the network-wide throughput while meeting user requirements. In this paper, we aim to address the problem of multi-region traffic scheduling, which seeks to deploy load balancing in LEO satellite networks. The contributions of this paper can be summarized as follows:

(i) We provide a traffic management architecture for LEO satellite networks. The architecture can realize multi-region traffic grooming in LEO satellite networks.

(ii) We propose a multi-network-region traffic cooperative scheduling (MR-TCS) algorithm. MR-TCS can improve the overall network throughput with different priorities.

(iii) We evaluate the proposed algorithm with existing load balance schemes. The results show that the proposed MR-TCS reduces the packet loss probability by 30% and improves the resource utilization ratio by 3.69%.

The rest of this article is organized as follows. Section 2 describes the LEO satellite network's traffic management architecture. We then introduce the proposed MR-TCS in Section 3. The evaluation is performed and discussed in Section 4. Then, we discuss the conclusions.

2. Related work

Due to the unbalanced distribution of user terminals on the ground, the traffic distribution of LEO satellite networks is unbalanced, which could produce congestion and underutilization in satellite networks. Therefore, the load balancing problem has become an essential issue in LEO satellite networks and has attracted much attention in recent years. Many routing algorithms for load balancing in satellite networks have been proposed. Based on the location where the algorithms are performed, these load-balancing routing schemes can be classified into two types: centralized and distributed [18].

2.1 Centralized load balancing schemes

A centralized routing algorithm means a controller computes the network routes. This controller periodically collects the state of each link and periodically provides routing tables to each network node after routing computation. A state-aware and load-balanced (SALB) algorithm is proposed in [19]. The main idea of SALB is to dynamically set the weight of the queuing delay based on the link state. It uses an efficient shortest path tree to reduce routing overhead and improve network throughput. A minimum traffic maximum residual routing scheme is proposed in [20], which ensures the load-balancing performance of satellite networks by distributing the service load over the shortest path alternative. A network coding based multipath cooperative routing scheme is proposed in [14], which dynamically uses Dijkstra's algorithm to transmit data along multiple links disjoint paths. Also, the scheme performs continuous data transmission using network coding and a no-stop-wait acknowledge mechanism, which results in good throughput and low end-to-end delay. The above works mainly focus on designing centralized routing schemes for LEO satellite networks. However, it can significantly affect propagation delays and overheads without considering the distance of time-varying ISLs in satellite networks.

2.2 Distributed load balancing schemes

The main idea of the distributed schemes is that each satellite in the network independently decides on the best next hop for forwarding packets. Compared with centralized and source-based load-balancing schemes, distributed load-balancing schemes can respond quickly to traffic changes. Ekici et al. [21] proposed a fully distributed datagram routing method to compute the minimum propagation delay path with small computational and storage overhead. Routing decisions are made based on local information, and no control messages are required, so routing is relatively static. Henderson et al. [22], Soret et al. [23], and Liu et al. [24] proposed comparable distributed routing computation methods. Henderson et al. [22] considered geographic information while considering routing paths and minimized the remaining distance. Soret et al. [23] tried to use as few inter-plane links as possible for routing. Liu et al. [24] used probabilistic routing to reduce the computational complexity. The distributed load-balancing schemes mentioned above solve the load-balancing problem of the network to some extent. However, distributed schemes have bottlenecks in obtaining real-time state data across the network. These schemes are prone to cause localized congestion in the overall network.

To address these problems, we consider grooming overflowed traffic into cold areas for transmission, which balances the load in the routing computing stage. For the traffic hotspot areas, the proposed algorithm reasonably diverts the traffic to improve the overall service capability of the network. While solving the load imbalance problem of the LEO satellite network, our method can discriminate between traffic flows with different QoS and perform multi-level transmission approaches.

3. System model

We first introduce the system model of the time-sensitive system. Then we discuss the queuing scheduling mechanism for LEO satellite networks.

3.1 Network architecture

The existing LEO satellite networks include a constellation system of tens of thousands of satellites. Each satellite can serve as an access node for satellite-terrestrial services and a switching node for inter-satellite traffic. Therefore, STLs and multiple ISLs are important components of the existing satellite networks. For a large-scale LEO satellite network, we need to divide it into multiple regions for management. In our proposed satellite network traffic management architecture, as shown in Fig. 1, the software-defined network (SDN) architecture divides the satellite network into data, control, and application planes. The data plane differs from the traditional terrestrial network and contains terrestrial access nodes, satellite network nodes, ISLs, and STLs. The control plane contains two types of controllers: the inter-domain controller within the satellite network domain, responsible for the management and configuration of the regional network nodes, and the network master controller, which is responsible for the interaction with the inter-domain controllers. The master controller can be deployed near the ground-based gateway stations to configure and distribute inter-satellite routing rules and traffic balancing rules. The application plane contains resource management and scheduling functions related to network services. The application plane unifies and controls the network's computing, storage, and bandwidth resources and allocates and monitors the traffic for differentiated services.

For traffic scheduling of services, our proposed architecture presents a management scheme. First, the control plane is aware of the distribution of network traffic and the network state. When a service request occurs, the controller sends the request message to the control plane for resource scheduling. When inter-domain resource collaborative scheduling is required, the intra-domain controller reports to the network master controller. After the network master controller receives the request, it assigns the task requirements to each manager through the message parser. Then, the corresponding service managers, including routing rules, service identification, load balancing, etc., calculate the forwarding actions of the services and finally form configurations to be distributed to the respective intra-domain controllers. Each node in the data plane is also responsible for ISL switching and fault management in regular forwarding, ensuring that all services in each LEO satellite network are allocated on-demand resources. In this architecture, the distribution of service traffic requirements should be monitored in the data plane, collaborating with the management and configuration in the control plane and application plane. The planning of services by the inter-domain master controller can be provided with the optimal resource-traffic match. With the proposed traffic management architecture, the service forwarding rules of the network can be dynamically configured and flexibly managed.

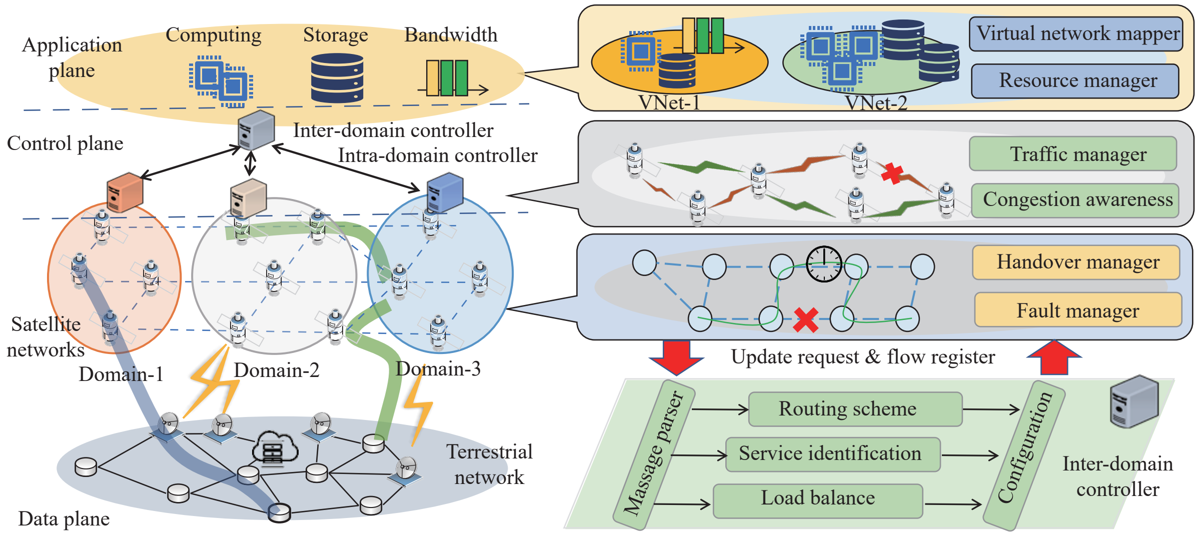


Fig. 1 Satellite-terrestrial convergence network architecture

3.2 Network model

An LEO satellite constellation is modeled as a graph $G=(V,E)$. V is the set of the satellites while E is the set of ISLs between satellites. There are N satellites in the constellation. The ISLs between any two neighboring satellites include intra-plane ISLs where the two satellites are in the same orbit plane and inter-plane ISLs for satellites in the neighboring orbit planes. Every satellite has two intra-plane ISLs and two inter-plane ISLs. The intra-plane ISL connects neighboring satellites between its preceding and succeeding satellites in the same orbit. The inter-plane ISL connects neighboring satellites in two neighboring orbits. In our model, each satellite has four ISLs. Each ISL has a bandwidth of B_s. The LEO satellite network forms a mesh topology. The number of STLs and user terminals in the ground covered by each satellite is random. The bandwidth of each user is fixed at B_t. The satellite-terrestrial services are randomly generated, and we limit the number of user terminals to a maximum of N_t. Therefore, the bandwidth of the satellite-terrestrial traffic received by a satellite is at a maximum of $Nt\cdot Bt$.

Based on the above constellation configuration, the topology of the network model is shown in Fig. 2. The topology is an $m\cdot n$ rectangle. The $S_{m,n}$ indicates a total of m orbits in the constellation, with n satellites in each orbit. Meanwhile, the ISL within the orbits remains stable because the relative positions of the satellites within the orbits remain the same in the inclined orbits; the distance and angle of the adjacent inter-orbit satellites change, but the relative topology remains the same. The satellites are always adjacent between the orbits, and the ISL may break at the pole areas. Therefore, the ISLs of this topology change with time.

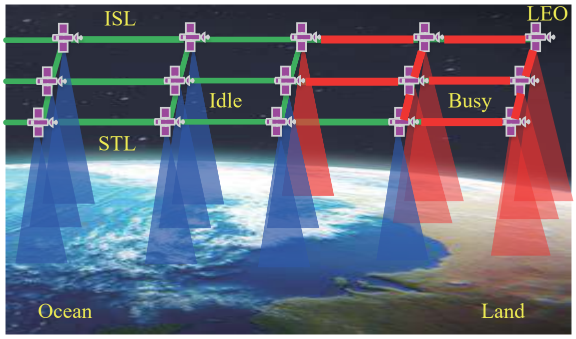


Fig. 2 Example of traffic scenarios for LEO satellite networks

In this model, we start with the physical layer to analyse the characteristics of the communication. The maximum achievable data rate for a given ISL can be calculated using the Shannon-Hartley theorem

$$C=B_s\log_2\left(1+\frac{S}{N}\right) \tag{1}$$

where C is the channel capacity in bits per second, B is the bandwidth in Hertz (Hz), and S/N is the signal-tonoise ratio.

The delay of a packet transmitted between two LEO satellites can be approximated using the following equation:

$$T_{\text{prop}} = \frac{2h}{\text{c}} \sqrt{1-e^2} \sin^{-1}\left(\frac{\cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 \cos(\Delta\phi)}{\sqrt{1-e^2}}\right) \tag{2}$$

where h is the altitude of the LEO satellite orbit, c is the speed of light, e is the eccentricity of the orbit, θ_1 and θ_2 are the latitudes of the two satellites, and $\Delta\phi$ is the difference in longitudes between the two satellites.

The end-to-end delay of a packet in the LEO satellite network can be computed as the sum of the ISL delay, the propagation delay, and the processing and queuing delays at the LEO satellite and ground station:

$$T_{\text{e2e}} = T_{\text{prop}} + T_{\text{processing}} + T_{\text{queuing}} \tag{3}$$

where $T_{\text{processing}}$ and T_{queuing} depend on the specific hardware and software configurations of the LEO satellite and ground station. The equation shows that the end-to-end delay is related to the transmission distance and the number of hops forwarded. The routing matrix used in a distributed routing protocol in a LEO satellite network can be calculated using the following equation:

$$M_{i,j} = \eta D_{i,j} + (1-\eta) L_{i,j} \tag{4}$$

where $D_{i,j}$ is the distance between satellite i and satellite j, $L_{i,j}$ is the link cost between satellite i and satellite j, and η is a weight factor that balances the importance of distance and link cost in the routing decision. These equations can be used to analyse and optimize the performance of traffic forwarding and routing in LEO satellite networks.

4. Problem formalization and motivation

Based on the description of the network model above, the routing of network traffic is not only related to the network topology but also impacted by the current traffic distribution and network states. Especially in large-scale LEO satellite networks, the impact of burst traffic is more evident due to individual satellite's limited computing and transmission resources. However, due to the characteristics of the distribution of the Earth's landscape, the traffic distribution is quite variable. If the satellite's STLs cover the oceans, the traffic volume in the coverage area is relatively small; in contrast, the satellite's STLs covering the major cities on land have a high traffic load. As shown in Fig. 2, the coverage area of satellites is indicated by blue for oceans and red for the land. The network in the right part is congested. Due to the difference in the geographical distribution of user terminals, the LEO satellite network's resources are hard to fully utilize. Therefore, improving the utilization rate of the LEO networks becomes an essential goal for future network development.

We improve the resource utilization of the overall LEO satellite network by grooming the traffic from land areas to ocean areas. Due to the geographical differences in the distribution of user terminals, it is crucial to improve the resource utilization of the network in coldspot areas. In existing studies, traffic is generally forwarded by the shortest-path-first (SPF) algorithm, which takes the network congestion status or hop count as optimization objectives. However, it is difficult for the existing methods to vary the manner of forwarding of services with different QoS requirements. Especially for time-sensitive traffic, the existing traffic scheduling methods lack differentiated forwarding capability. Focusing on this problem, we try to provide a multi-network-region collaborative scheduling algorithm for different latency requirements. In our model, the forwarding range of the services is identified first, and the services with QoSs are limited to different range scales for transmission. Thus, the resource utilization of the whole network is improved.

5. Multi-region cooperative traffic scheduling algorithm

In this section, we first introduce how to measure the queuing delay, then present a region-locking mechanism, and finally propose the multi-region collaboration SALB routing algorithm.

5.1 Link state and queuing delay

The queuing delay depends on the link state, which captures the traffic load over an ISL. We define the link occupancy ratio $R_{(m,n),(i,j)}$ of an ISL from satellite $S_{m,n}$ to $S_{i,j}$ based on the buffer occupancy

$$R_{\{(m,n),(i,j)\}} = \frac{Q_{\{(m,n),(i,j)\}}}{Q_{\text{size}}} \tag{5}$$

where Q_{size} denotes the buffer size of the ISL, and $Q_{\{(m,n),(i,j)\}}$ denotes the buffer size occupied by the packets of the queue. This paper assumes all satellites have the same buffer size and capacity. The maximum queuing delay of the ISL for the current queue's traffic is

$$L_{\{(m,n),(i,j)\}} = \frac{Q_{\{(m,n),(i,j)\}}}{C}. \tag{6}$$

To facilitate queue management, we transform the buffer occupancy ratio from continuous values to discrete values

$$R_{\{(m,n),(i,j)\}} = \lfloor n \cdot \frac{Q_{\{(m,n),(i,j)\}}}{Q_{\text{size}}} \rfloor \tag{7}$$

where $R_{(m,n),(i,j)}$ takes values in $\{1, 2, \cdots, n\}$.

Specifically, if $R_q(l_{i,j})$ falls into the range $\left[\dfrac{k-1}{n}, \dfrac{k}{n}\right)$, we set the link state $l_{i,j}$ as $(k-1)$.

However, the queue latency does not fully reflect the carrying capacity of the buffer. Since the length of the queue is limited, there may be packet losses when the queue is fully occupied. Therefore, the impact of traffic on the queue should consider both queue delay and packet loss.

Therefore, we introduce the possibility of packet loss into the calculation of $R_{\{(m,n),(i,j)\}}$. The packet loss probability factors include the buffer size, round-trip time (RTT) delay, and the number of flows. We calculate a factor θ by the following equation to characterize the impact of packet loss:

$$\theta = \frac{T_{\text{prop}} \cdot C}{(Q_{\text{size}} - Q_{\{(m,n),(i,j)\}})}. \tag{8}$$

Although the queuing delay increases linearly as $R_{\{(m,n),(i,j)\}}$ increases, inserting the packet incurs a high risk of losing packets when the buffer becomes almost complete in the last few states. To migrate flows to low-load links, we combine packet loss and latency into computing the weight of an ISL. It is inversely proportional to the possible ratio of packet loss to the idle buffer size and directly proportional to the RTT latency.

To lower the weight of links with low $Q_{\{(m,n),(i,j)\}}$, we improve $L_{i,j}$ in (6) with the following weight:

$$L_{(m,n),(i,j)} = R_{\{(m,n),(i,j)\}} \cdot \theta \tag{9}$$

where $\theta(\theta > 0)$ is an adjustment factor of the ISL to balance the impacts of distance and weight. The basic idea of this model is to consider both latency and reliability because it is desirable to expand the routing region while forwarding the multi-level traffic.

Based on the value of $L_{(m,n),(i,j)}$, we can correctly react to the impact of the link state on the network performance. Therefore, the routing strategy can be determined based on the values of the ISL weights. In addition, the above equation links the buffer occupancy to multiple performances to accurately characterize the traffic. Hence, we can adapt the routing scheme according to the coldspot and hotspot areas based on $L_{(m,n),(i,j)}$.

5.2 Routing region division mechanism

According to the LEO satellite network model, we must collect the satellite's node and link states in order to calculate the routing. Since the satellites in the busy state are clustered, the busy regions are only related to the geographical location, which remains stable in geographical distribution. Therefore, we divide the satellite network into hotspot, hotspot radiating, and coldspot areas, as shown in Fig. 3. We first calculate the location of the busy satellites and lock the hotspot area, which is indicated by red. The nodes in the yellow area are the hotspot radiating area, which still has more traffic flow than the coldspot area. Then, most of the ocean is a coldspot. The coldspot area has less traffic than other areas and has a relatively sufficient transmission bandwidth. Among them, the division of the region type is determined by the long-term traffic statistics. For example, we set two parameters α and β $(\alpha \geqslant \beta)$. If the average traffic of the satellite is greater than α, it is in a hotspot area. The satellite is in a cold area if the average satellite traffic is less than β.

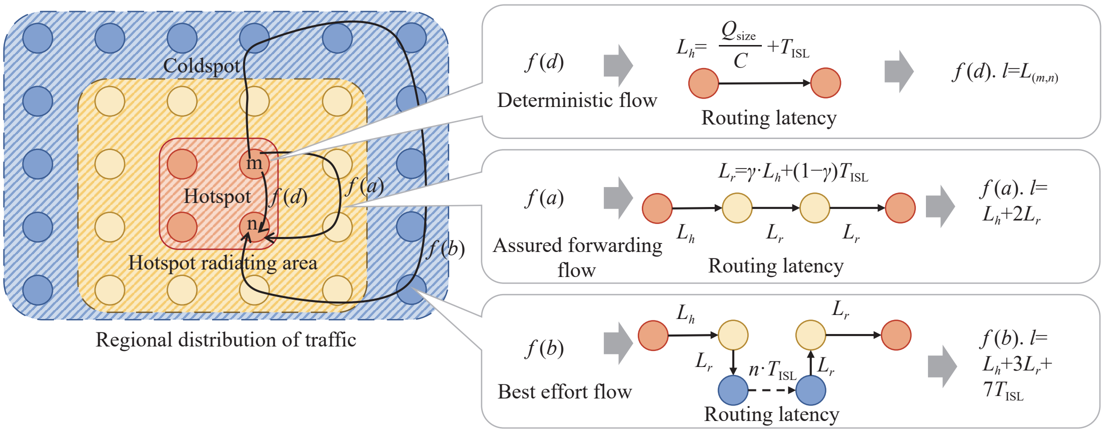


Fig. 3 Example for routing area division mechanism

After the division into regions, we can periodically adjust the direction of migration of the hotspot region in the satellite network according to the geographical location of the satellite. As shown in Fig. 3, we categorize the traffic flow into deterministic flow (DF), assured forwarding (AF), and best-effort (BE) traffic. For deterministic traffic, the routing algorithm of the SPF is used. Since propagation delay accounts for most of the transmission latency in satellite networks, the SPF algorithm ensures the shortest latency for the traffic. In hotspot areas, the maximum latency of each hop is given by

$$L_h = \frac{Q_{\text{size}}}{C} + T_{\text{ISL}} \tag{10}$$

where L_h considers the maximum queuing delay and propagation delay. For assured forwarding flow, both the latency and the reliability of forwarding are considered. Therefore, the network forwards assured traffic through hotspots and radiating areas. Like deterministic flow, the distance matrix is generated for the two areas before routing, and then the routing algorithm calculates the paths. In the process of calculating the routing, we set a coefficient γ to simplify the calculation of radiating areas to avoid frequent state collection. Then the latency for each hop in the hotspot radiating areas is

$$L_r = \gamma \cdot L_h + (1-\gamma) \cdot T_{\text{ISL}} \tag{11}$$

where γ denotes the queuing delay ratio of hotspot area to hotspot radiating area. The γ is an adjustable parameter resulting from a long-term statistical flow. The proposed model determines γ based on the shortest distance to other areas. The γ of a node near a hotspot area is more significant than that near a coldspot area. If the node in a radiating area connects both hotspot and coldspot areas, γ is set to 0.5. As shown in Fig. 3, the traffic flow $f(a)$ has a latency of $(L_h + 2L_r)$ after three hops. It can be forwarded through the coldspot region in a larger spatial scale range for best-effort traffic. Since there are fewer flows in the coldspot area, the propagation delay is used to characterize the latency of each hop. The end-to-end latency of traffic flow $f(b)$ after passing through 11 hops, as shown in Fig. 3, is $(L_h + 3L_r + 7T_{\text{ISL}})$. This approach increases the forwarding latency of low-level flows but dramatically improves the overall network throughput.

5.3 Multi-region cooperative scheduling algorithm

After the determination of the regions of the LEO satellite network is completed, the routing algorithm needs to consider how to schedule traffic to achieve traffic balancing in different regions. First, there is an SDN controller in the LEO satellite network to calculate the routes. This controller can monitor the traffic in the network and obtain ISL status information. Therefore, when a new traffic flow appears, the source satellite node sends a route computation request to the controller via a packet in message on the southbound interface. The controller gets the request, calculates the service distribution and constellation configuration by the proposed traffic scheduling and routing algorithm, and then implements the service flow transmission in the whole network. The controller maps the computed routes to each satellite's lookup table configuration information. Finally, the controller sends the configuration messages to the satellites to realize traffic forwarding.

Moreover, we design the routing algorithm to realize the scheduling between multiple regions. The routing algorithm is shown in Algorithm 1. Firstly, the controller converts the network traffic statistics into a queuing delay matrix ($\boldsymbol{L}$) and then calculates the propagation delays of the ISL links by constellation states ($\boldsymbol{D}$). After completing the computation of the two matrices $\boldsymbol{L}$ and $\boldsymbol{D}$, we combine them into the routing matrix $\boldsymbol{M}$. Each entry of $\boldsymbol{M}$ denotes the weight of a link in the routing computation. However, our routing method needs to specify the spatial extent of the routing calculation. Therefore, based on the routing matrix, we need to calculate the type of area to which each satellite belongs. We define two thresholds, α and β ($\alpha > \beta$). When the statistical value of a link is more significant than α, it is in a hotspot region. When it is at the edge of the hotspot area, the node is in the hotspot radiation area. The node is in the coldspot region when the statistic value is less than β.

Algorithm 1: Multi-region cooperative scheduling algorithm

Input: The traffic matrix set $\boldsymbol{M}_t$, the distance matrix set $\boldsymbol{D}_t$, the maximum ISL capacity C, the request flow f^*, area division coefficient α, β and weighting factor η

Output: The routing path P of f^*

```
Initialization: $\boldsymbol{M}_t$, $\boldsymbol{D}_t$
Updating queuing delay matrix $\boldsymbol{L}_{i,j}$:
for $\forall\, i,j \in \boldsymbol{L}_{i,j}$ do
    Calculate $\boldsymbol{L}_{i,j}$ according to (9).
endfor
Updating propagation delay for each ISL, as $\boldsymbol{D}_{i,j}$:
for $\forall\, i,j \in \boldsymbol{D}_{t(i,j)}$ do
    Calculate $\boldsymbol{D}_{i,j}$ for each $\boldsymbol{D}_{t(i,j)}$ according to (2).
endfor
Routing matrix: $\boldsymbol{M} \leftarrow \eta\boldsymbol{D} + (1-\eta)\boldsymbol{L}$
for $\forall\, i,j \in \boldsymbol{M}_{i,j}$ do
    if $\boldsymbol{M}_{i,j} \geqslant \alpha$ then
        Set $\boldsymbol{M}_{i,j}$ as hotspot node.
    endif
for $\forall\, i,j \in \boldsymbol{M}_{i,j}$ do
    if $\boldsymbol{M}_{i,j}$ adjacent to hotspot area then
        Set $\boldsymbol{M}_{i,j}$ as hotspot radiating node.
    else
        if $\boldsymbol{M}_{i,j} \geqslant \beta$ then
            Set $\boldsymbol{M}_{i,j}$ as hotspot radiating node.
```

```
        else
            Set $M_{i,j}$ as coldspot node.
        endif
    endif
endfor
Calculate weight of $M$ as the node type.
Lock the routing area in $M$
Calculate the path by SPF as $P$.
return $P$
```

After completing the region classification, $\boldsymbol{M}$ adjusts the weights to lock the routing regions. The satellite is categorized as a member of one of the sets S_h, S_r and S_c, denoting the hotspot area, radiating area and coldspot area, respectively. The proposed algorithm has different $\boldsymbol{M}$s for the three types of traffic. Since the DF traffic should be guaranteed by optimal routing, the routing algorithm uses global information to select the shortest path. Then, $\boldsymbol{M}$ is given by

$$M_{i,j}^{\mathrm{DF}} = D_{i,j} + L_{i,j},\ \ \forall i,j. \tag{12}$$

For AF traffic, the proposed algorithm needs to ensure that it can be forwarded efficiently, but without time-sensitive characteristics. Therefore, it can be forwarded through the radiating area and we represent its routing matrix $\boldsymbol{M}$ by the following equation:

$$M_{i,j}^{\mathrm{AF}} = \begin{cases} \tau,\ \ s_{i,j} \in S_h \\ D_{i,j} + L_{i,j},\ \ s_{i,j} \in S_r \\ L_{i,j},\ \ s_{i,j} \in S_c \end{cases} \tag{13}$$

where τ is a relatively large number compared to $D_{i,j}$ and $L_{i,j}$. In routing calculations, τ is prioritized over a range of values in $D_{i,j}$ and $L_{i,j}$ ($\tau \gg D_{i,j}, L_{i,j}$).

For BE flows, the proposed algorithm transmits as many flows as possible through the coldspot region. Meanwhile, the weights of hotspot regions are higher than the weights of radiating regions. Therefore, we define two constants to characterize the weights of the hotspot region and radiating region: τ and κ ($\kappa \gg \tau$). Hence, $\boldsymbol{M}$ is

$$M_{i,j}^{\mathrm{BE}} = \begin{cases} \kappa,\ \ s_{i,j} \in S_h \\ \tau,\ \ s_{i,j} \in S_r \\ L_{i,j},\ \ s_{i,j} \in S_c \end{cases}. \tag{14}$$

After determining $\boldsymbol{M}$, the routing calculation is completed by the SPF. The proposed algorithm is similar to the core computational process of the SPF algorithm. Therefore, the algorithmic complexity of MR-TCS is $O(n^2)$, where n is the number of nodes in the network domain. After the algorithm finishes calculating the routes, it outputs the available path P. Then, the controller assembles the routing configuration information for the satellite corresponding to P and sends it down to each satellite to realize traffic forwarding. In this algorithm, we forward the traffic in the hotspot area through the hotspot radiating area and the coldspot area to avoid traffic congestion in the hotspot area.

6. Performance evaluations

6.1 System setting

In this section, we build a testbed for the proposed algorithm and other typical traffic scheduling techniques in LEO satellite networks. To evaluate the routing region division mechanism, we randomly generate traffic load for each satellite in mesh networks and define the region type for each node by our proposed mechanism, which is dedicated to validating the latency and hops of routing paths. Then we simulate an 8*8 Walker constellation to observe the packet transmission performance. We adopt some typical algorithms for the comparison, including SPF, ant colony optimization (ACO) and SALB [19].

The SPF algorithm uses a greedy routing algorithm, which always finds the path with the shortest distance, thus the path with the smallest propagation delay. ACO is a kind of heuristic algorithm with excellent performance for routing. SALB can quantitatively estimate link states and dynamically adjust the weight of queuing delay, and is claimed to be better than typical algorithms. We build a packet-level testbed for the LEO satellite networks by OMNET++.

The Walker constellation is organized as a mesh network. The satellite in the constellation has four ISLs, including two inter-orbit and two intra-orbit ISLs.

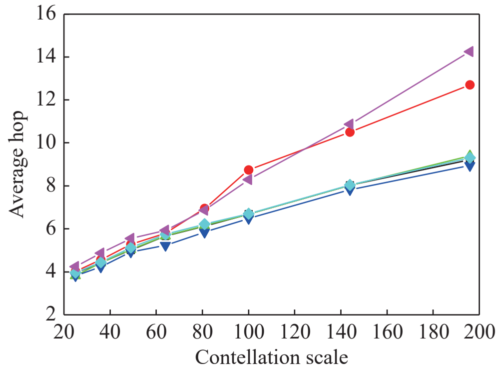
The traffic flow is generated by an "ON–OFF" model with lengths following a Pareto distribution. The size of the data packets is between 64 B and 1 518 B. The size of traffic flows ranges from 64 B to 50 MB and follows a Pareto distribution with parameters 64, 0.9, and 3 854, which ensures that 80% of the streams are less than 1 MB. We change the normalized load's size by controlling the OFF period's length. The packet loss probability is a statistical result of packets across the network. Each satellite node generates 100 k packets, and at the end of the simulation, the proportion of dropped packets to all packets is counted. In the simulation, we only consider packet loss caused by buffer overflow. The packet size follows a bimodal distribution as described in [25]. The source and destination of the traffic flows are random among all satellites, where each flow has the same destination. At the same time, the service QoS of traffic is categorized into three types: deterministic traffic, assured forwarding traffic, and best-effort traffic. The three types are randomly generated with a volume ratio of 3:3:4. We inject the traffic into the network and collect the statistics to calculate the performance. In the simulation, each ISL has a buffer size of 1 MB, and we set the ISL length to vary from 2 000 km to 2 500 km. With a balanced load scenario, we set different nodes under different load cases. We use a normalized load pattern and set the half of the nodes in the middle of the topology (rounded up) to high load (0.8) and the rest of the satellite traffic to low load (0.2).

6.2 Routing and latency

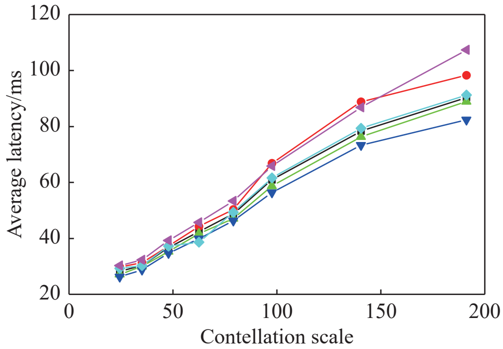
We first validate the average hop count and average latency performance of the network traffic. We conduct the routing algorithm performance tests in constellations of different scales. The paths are then assigned to each traffic flow by routing algorithms that count the average hop count and average delay of the packets. Except for ACO, which uses a heuristic algorithm, all other algorithms obtain routes by the SPF method, and the calculation of the weights for the routing algorithms differs.

Fig. 4(a) shows the average hop count of packets with different algorithms and constellation scales. Compared with the heuristic algorithm ACO, the SPF algorithm can reduce the average hop count of traffic. Due to the large spatial range of the satellite networks, the latency of the traffic flows can be significantly reduced. Meanwhile, our proposed MR-TCS approach forwards the traffic of different types in different ways. The high-priority services can maintain almost the same average hop count as the conventional method. However, the low-priority service (BE) introduces a large average hop count due to using a larger network area to evacuate the traffic.

Fig. 4(b) shows the average end-to-end latency under different algorithms. Compared to typical algorithms, our proposed MR-TCS can reduce the average experiment of DF services by 7.39%. However, it greatly increases the latency of low-priority services due to the traffic grooming to the coldspot region for forwarding. In satellite networks, in addition to propagation delay, queuing delay still affects end-to-end delay. MR-TCS can reduce the queuing delay of DF traffic by grooming low-priority traffic to other paths, thus maintaining the best end-to-end delay performance for DF traffic. Meanwhile, the difference in average delay is small compared to the difference in average hop count. The effect of traffic grooming reduces the queuing delay of traffic in the transmitter buffer. Therefore, the proposed approach is of great significance for securing high-priority traffic.



(a) Average hops vs. constellation scales for all flows with a buffer size of 1 MB and an ISL range of 2 000 km



(b) Average latency vs. constellation scales for all flows with a buffer size of 1 MB and an ISL range of 2 000 km

–■–: SPF; –●–: ACO; –▲–: SALB; –▼–: MR-TCS(DF); –◆–: MR-TCS(AF); –◀–: MR-TCS(BE).

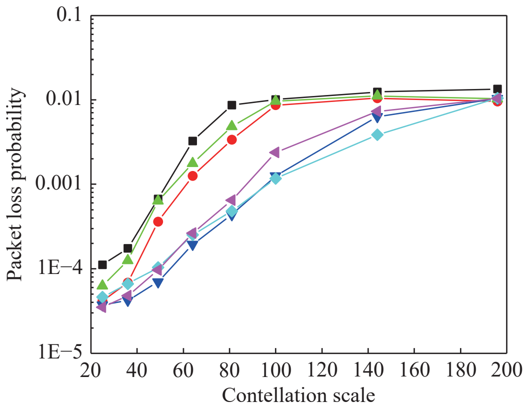
Fig. 4 Routing comparison of different algorithms with different network scales

6.3 Packet loss

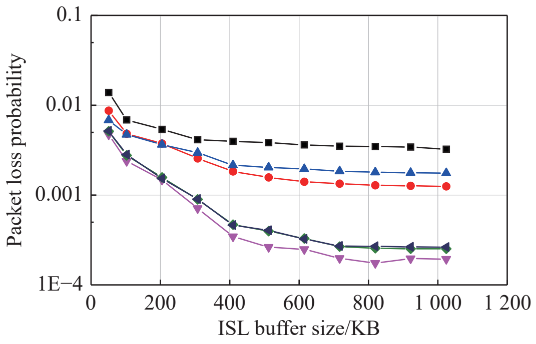
The packet loss probability is an important network performance metric and using a different routing algorithm can significantly affect the packet loss. We counted the packet loss probability under different constellations and buffer sizes to validate the performance of different routing algorithms with unbalanced load cases. In the resultant statistics, 10^5 packets are generated by each user terminal, and the packet loss rate is counted after forwarding through the LEO satellite network. Among them, we adopt the retransmission mechanism in the simulation. If a packet loss occurs during the forwarding, the source node will resend the packet only once.

Fig. 5(a) shows the network packet loss probability under different constellation scales. In the described load case, the larger the topology, the higher the packet loss probability. Meanwhile, the proposed MR-TCS shows the lowest packet loss probability. The advantage of the proposed algorithm is more prominent, especially in the case of a large topology. Furthermore, the proposed algorithm's packet loss performance for all three service types is significantly improved. Due to grooming low-priority services to the coldspot area for forwarding, MR-TCS can significantly reduce the packet loss probability of low-priority services while providing more bandwidth resources for high-priority services.

Fig. 5(b) shows the packet loss probability performance under different ISL buffer sizes. The packet loss probability of different routing algorithms decreases as the buffer size increases. Moreover, the packet loss probability increases significantly in the case of a buffer size below 400 KB; the correlation between the packet loss probability remains stable when the buffer size is higher than 400 KB. The MR-TCS algorithm proposed in the paper has the lowest packet loss rate. Compared with the traditional algorithm, MR-TCS reduces the packet loss rate by 30% under all load cases. Therefore, the proposed algorithm maintains the best packet loss performance for different constellations and buffer sizes.



(a) Packet loss probability vs. load with ISL queue size of 153.6 KB



(b) Packet loss probability vs. ISL queue size at loads of 0.4 and 0.9

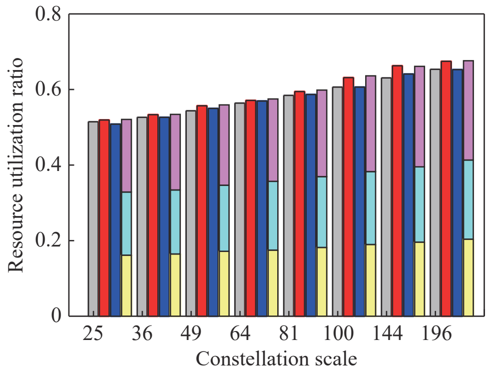
—■—: SPF; —●—: ACO; —▲—: SALB; —▼—: MR-TCS(DF);
—◆—: MR-TCS(AF); —◀—: MR-TCS(BE).

Fig. 5 Packet loss performance of different algorithms

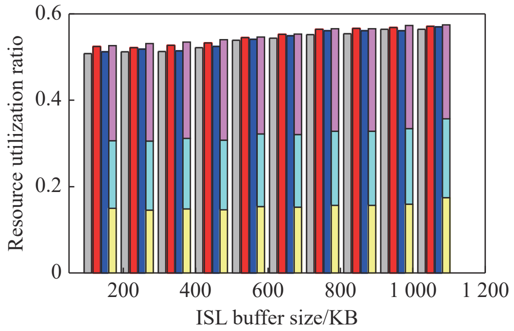
6.4 Resource utilization

Network resource utilization is an essential indicator of network performance. Whether the bandwidth resources can be effectively utilized under different routing algorithms indicates the carrying capacity of the LEO satellite network. We define the resource utilization ratio as the ratio of the entire network’s occupied bandwidth resources to the total bandwidth resources. Due to the LEO satellite network’s highly dynamic and unbalanced load, it is difficult to improve network resource utilization effectively. Therefore, we must improve resource utilization in the coldspot area to improve the network performance overall.

Fig. 6(a) shows the network resource utilization ratios under different constellation scales. As the constellation scale increases, the resource utilization ratio improves. On the one hand, because the distribution of traffic in large-scale networks is more dispersed, more bandwidth resources are effectively utilized; on the other hand, the average number of hops for service transmission increases, resulting in more resources being required for service transmission. It is the core objective of the routing algorithm to be able to fully utilize the resources of the whole network and reduce the packet loss.



(a) End-to-end latency vs. load with ISL queue size of 460.8 KB,TSN updating period of 250 μs



(b) End-to-end latency vs. ISL queue size with an TSN updating period of 250 μs at a load of 0.6

▨: MR-TCS(BE); ▨: MR-TCS(AF); ▨: MR-TCS(DF);
▨: SPF; ▨: SALB; ▨: ACO.

Fig. 6 End-to-end latency of different algorithms

From Fig. 6(a), we can see that our proposed MR-TCS maintains the highest resource utilization ratio. The resource utilization of the ACO algorithm is almost equal to the proposed MR-TCS. However, the ACO algorithm requires a higher average hop count. Therefore, the good resource utilization ratio of ACO comes at the expense of a significant amount of latency. Furthermore, using a larger buffer can increase resource utilization, and the advantage of MR-TCS is more evident with a large buffer. The BE service occupies more bandwidth resources among the three service types because a more distant path is used. Compared with SALB, the resource utilization ratio of the proposed algorithm is improved by 3.69%. Therefore, the proposed algorithm in this paper can improve resource utilization under different constellation scales.

Fig. 6(b) shows the resource utilization with different ISL buffer sizes. It can be seen that resource utilization improves as the buffer increases. The algorithm proposed in this paper maintains the highest bandwidth resource utilization ratio because the multi-region collaboration scheduling can improve the carrying capacity of the network. Although the improvement in the resource utilization ratio is insignificant, the small difference in the average resource utilization ratio across the network indicates many services. At a buffer size of 200 KB, the proposed algorithm improves the performance over the SALB algorithm by 0.9%.

7. Conclusions

The paper investigates traffic scheduling in a large-scale LEO satellite network under unbalanced traffic load conditions. An MR-TCS scheme is proposed so as to address the characteristics of an LEO satellite network with large regional differences in geographic locations. Firstly, the method divides satellite nodes into hotspot regions, hotspot radiation regions, and coldspot regions. Then, the scheme categorizes the existing services into three types and grooms the low-priority services to the coldspot region for forwarding to improve the network resource utilization. In light of simulations at different constellation scales, it has been found that the proposed method can reduce the packet loss probability and improve the network resource utilization ratio. Compared with conventional algorithms, MR-TCS reduces the packet loss probability by 30% and improves the resource utilization ratio by 3.69%. This approach can provide a technical basis for constructing LEO satellite constellations.

References

[1] WANG R B, KISHK M A, ALOUINI M-S. Ultra-dense LEO satellite-based communication systems: a novel modeling technique. IEEE Communications Magazine, 2022, 60(4): 25–31.

[2] ZHEN L, BASHIR A K, YU K, et al. Energy-efficient random access for LEO satellite-assisted 6G internet of remote things. IEEE Internet of Things Journal, 2021, 8(7): 5114–5128.

[3] GIAMBENE G, KOTA S, PILLAI P. Satellite-5G integration: a network perspective. IEEE Trans. on Networking, 2018, 32(5): 25–31.

[4] LIU S C, GAO Z, WU Y P, et al. LEO satellite constellations for 5G and beyond: how will they reshape vertical domains? IEEE Communications Magazine, 2021, 59(7): 30−36.

[5] LIU W, TAO Y, LIU L. Load-balancing routing algorithm based on segment routing for traffic return in LEO satellite networks. IEEE Access, 2019, 7: 112044–112053

[6] CAINI C, FIRRINCIELI R. Application of contact graph routing to LEO satellite DTN communications. Proc. of the IEEE International Conference on Communications , 2012: 3301−3305.

[7] GOTO D, SHIBAYAMA H, YAMASHITA F, et al. LEO-MIMO satellite systems for high capacity transmission. Proc. of the IEEE Global Communications Conference, 2018. DOI: 10.1109/GLOCOM.2018.8647694.

[8] SHI K Y, LI H Y, SUO L. Temporal graph-based energy-limited maxflow routing over satellite networks. Proc. of the IFIP Networking Conference, 2021. DOI: 10.23919/IFIPNetworking52078.2021.9472794.

[9] GAO Z H, GUO Q, NA Z Y. Novel optimized routing algorithm for LEO satellite IP networks. Journal of Systems Engineering and Electronics, 2011, 22(6): 917–925.

[10] MARCANO N J H, DIEZ L, CALVO R A, et al. On the queuing delay of time-varying channels in low earth orbit satellite constellations. IEEE Access, 2021, 9: 87378–87390.

[11] WANG J, PHAM K. Design of nonlinear control for active queue management in TCP satellite communication networks. Proc. of the IEEE Aerospace Conference, 2020. DOI: 10.1109/AERO47225.2020.9172560.

[12] DENG X, CHANG L, ZENG S Y, et al. Distance-based backpressure routing for load-balancing LEO satellite networks. IEEE Trans. on Vehicular Technology, 2023, 72(1): 1240–1253.

[13] CUI Y, YEH E M, LIU R. Enhancing the delay performance of dynamic backpressure algorithms. IEEE/ACM Trans. on Networking, 2016, 24(2): 954–967.

[14] TANG F L, ZHANG H T, YANG L T. Multipath cooperative routing with efficient acknowledgement for LEO satellite networks. IEEE Trans. on Mobile Computing, 2019, 18(1): 179–192.

[15] WANG F, YAO H P, HE W J, et al. Time-sensitive scheduling mechanism based on end-to-end collaborative latency tolerance for low-earth-orbit satellite networks. IEEE Trans. on Network Science and Engineering, 2023. DOI: 10.1109/TNSE.2023.3342938.

[16] NING Y X, ZHAO Y L, LI X, et al. Load balancing routing algorithm against inter-satellite link congestion in LEO satellite optical networks. Proc. of the Optical Fiber Communications Conference and Exhibition, 2022. DOI: doi.org/10.

1364/OFC.2022.W2A.22.

[17] NISHIYAMA H, KUDOH D, KATO N, et al. Load balancing and QoS provisioning based on congestion prediction for GEO/LEO hybrid satellite networks. Proceedings of the IEEE, 2011, 99(11): 1998–2007.

[18] WU Z F, HU G Y, JIN F L, et al. A novel routing design in the IP-based GEO/LEO hybrid satellite networks. International Journal of Satellite Communications and Networking, 2017, 35(3): 179–199.

[19] LI X, TANG F L, CHEN L, et al. A state-aware and load-balanced routing model for LEO satellite networks. Proc. of the IEEE Global Communications Conference, 2017. DOI: 10.1109/GLOCOM.2017.8254443.

[20] KUCUKATES R, ERSOY C. Minimum flow maximum residual routing in LEO satellite networks using routing set. Wireless Networks, 2008, 14: 501–517.

[21] EKICI E, AKYILDIZ I F, BENDER M D. A distributed routing algorithm for datagram traffic in LEO satellite networks. IEEE/ACM Trans. on Networking, 2001, 9(2): 137–147.

[22] HENDERSON T R, KATZ R H. On distributed, geographic-based packet routing for LEO satellite networks. Proc. of the IEEE Conference on Global Communications, 2000. DOI: 10.1109/GLOCOM.2000.891311.

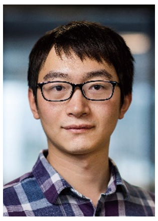
[23] SORET B, SMITH D. Autonomous routing for LEO satellite constellations with minimum use of inter-plane links. Proc. of the IEEE International Conference on Communications, 2019. DOI: 10.1109/ICC.2019.8761787.

[24] LIU X M, JIANG Z Q, LIU C H, et al. A low-complexity probabilistic routing algorithm for polar orbits satellite constellation networks. Proc. of the IEEE/CIC International Conference on Communications in China, 2015. DOI: 10.1109/ICCChina.2015.7448759.

[25] WANG F, YAO H P, ZHANG Q, et al. Dynamic distributed multi-path aided load balancing for optical data center networks. IEEE Trans. on Network and Service Management, 2022, 19(2): 991–1005.

Biographies

LI Chengxi was born in 2002. He is pursuing his Bachelor’s degree at Beijing University of Posts and Telecommunications from 2021, majoring in electronic information science and technology. His research interests include satellite networking, autonomous driving and intelligent control.
E-mail: lucas1558@126.com



WANG Fu was born in 1990. He received his B.S. degree in electronic information engineering from Hangzhou Dianzi University in 2013, and Ph.D. degree in electronic science and technology from Beijing University of Posts and Telecommunications (BUPT) in 2020. He was a Postdoctoral Researcher with Beijing Institute of Technology from 2020 to 2022 and is currently an associate research fellow with BUPT. His research interests include data center networks, satellite networks and intelligent optical networks.
E-mail: wangfu6672@gmail.com

YAN Wei was born in 1990. He received his B.S. degree in Taiyuan University of Technology in 2012, and Master’s degree in the 6th Research Institute of China Electronics Corporation in 2016, majoring in computer science and technology. He is Cyber Security Institute Technical Director and Master Tutor in the 6th Research Institute of China Electronics Corporation. Since 2022, he has been a senior engineer, as well as a Chinese Institute of Electronics Fellow. His research interests include industrial control system security, cloud computing security and cyberspace security.
E-mail: yanwei@ncse.com.cn

CUI Yansong was born in 1975. He received his Ph.D. degree from Beijing University of Posts and Telecommunications (BUPT) in 2005, majoring in microelectronics and solid-state electronics, and is now a full professor at BUPT. His research interests are artificial intelligence and multimedia communication.
E-mail: cuiys@bupt.edu.cn

FAN Xiaodong was born in 1986. He received his B.S. degree in Shenyang Aerospace University in 2009, and Master’s degree in electronic science and technology from Xi’an University of Electronic Technology in 2018. Science 2021, he has been a Deputy General Manager in China Satellite Network Share Operation Co., Ltd., as well as an engineer. His research interests include satellite Internet application.
E-mail: fxd5418102@163.com

ZHU Guangyu was born in 1981. He received his B.S. degree from Beijing Institute of Mechanical Technology in 2004, and Master’s degree in the 6th Research Institute of China Electronics Corporation in 2014, majoring in computer science and technology. He is the director of the Cyber Security Institute and Master Tutor in the 6th Research Institute of China Electronics Corporation. His research interests include network and information security.
E-mail: 34512553@qq.com

XIE Yanxi was born in 1986. She received her M.S. degree from Beijing University of Posts and Telecommunications-Emlyon Business School in France in 2021, majoring in business administration (Internet of Things). Her research interests include data center networks, information engineering and digital economics.
E-mail: 13671300662@163.com

YANG Lixin was born in 1987. He received his B.S. degree from Beijing Vocational College of Science and Technology in 2011, majoring in network application technology. His main research interests include smart building, smart park and smart data center.
E-mail: 15652287656@163.com

ZHOU Luming was born in 2002. He is pursuing his bachelor's degree at Beijing University of Posts and Telecommunications from 2021, majoring in optical information science and engineering. His research interests include artificial intelligence security and artificial intelligence related applications.
E-mail: 13509744323@163.com

ZHAO Ran was born in 2003. He is pursuing his bachelor's degree in Beijing University of Posts and Telecommunications from 2021, majoring in electronic science and technology. His research interests include data center networks, artificial intelligence security and artificial intelligence generated content.
E-mail: zhaoran03@bupt.edu.cn

WANG Ning was born in 2004. She is pursuing her bachelor's degree in Beijing University of Posts and Telecommunications from 2021, majoring in electronic science and technology. Her research interests include network and information security, artificial intelligence and multimedia communication.
E-mail: wangning042021@126.com