IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 29, 2018, accepted November 29, 2018, date of publication December 7, 2018,
date of current version January 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2885473

SDN-Based End-to-End Fragment-Aware Routing for Elastic Data Flows in LEO Satellite-Terrestrial Network

QIZE GUO[1], RENTAO GU[1], (Member, IEEE), TAO DONG[2], JIE YIN[2], ZHIHUI LIU[2], LIN BAI[3], AND YUEFENG JI[4], (Senior Member, IEEE)

[1]Beijing Laboratory of Advanced Information Network, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]State Key Laboratory of Space-Ground Integrated Information Technology, Beijing Institute of Satellite Information Engineering, Beijing 100095, China
[3]School of Economic Management, Beijing University of Posts and Telecommunications, Beijing 100876, China
[4]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Yuefeng Ji (jyf@bupt.edu.cn)

This work was supported in part by the Beijing Natural Science Foundation under Grant 4182040, in part by the National Natural Science Foundation of China under Grant 61771073, in part by the Open Research Fund of State Key Laboratory of Space-Ground Integrated Information Technology under Grant 2015_SGIIT_KFJJ_TX_03, in part by the CAST Fund for Distinguished Young Talents, in part by the Advanced Research Field Fund of Equipment Development, and in part by the CASC Scientific and Technological Innovative Research and Design Projects, China.

ABSTRACT The wide coverage of satellite networks and the large bandwidth of terrestrial networks have led to an increasing research on the integration of the two networks (ISTN) for complementary advantages. However, most researches on routing mainly focus on the internal routing of the satellite network. Due to the point-to-area coverage of the channel characteristics between the satellites and the ground stations, the heterogeneity of ISTN has increased, which makes that the routing algorithm of the traditional satellite networks cannot be applied to the end-to-end routing of ISTN. Meanwhile, the data flows with elastic quality of service attribute make the routing pre-assign and resource allocation in ISTN much more complicated, which has been rarely researched before. In this paper, we first describe the unified network architecture based on software-defined network and model the data flow. Then, based on the considerations of latency, capacity, wavelength fragmentation, and load balancing, a heuristic service-oriented path computation algorithm for elastic data flows is proposed for the complex heterogeneity of the ISTN. The simulation shows that, the end-to-end routing mechanism can reduce the blocking rate of the ISTN, and our proposed algorithm greatly reduces the wavelength fragmentation and bandwidth consumption, and has a better performance on load balancing, with a slight disadvantage in latency when the network load is high.

INDEX TERMS Low earth orbit satellites, quality of service, wavelength routing, optical fiber networks.

I. INTRODUCTION

As the satellite networks own the advantages of global coverage, mobility and scalability compare with the terrestrial networks which have developed with enormous transport capacity and low latency, the Integrated Satellite-Terrestrial Network (ISTN) has been proposed and brought into research for years with taking advantages of both networks to supply global services. As the two types of networks are heterogeneous, the related research on ISTN mainly focused on the centralized control [1]–[3], protocol interoperability [4] and some new scenarios [5]–[7], with little research explicitly addressing the routing problem for data flows in ISTN.

In order to improve the transport efficiency of the data flows, the IP/MPLS based on-board forwarding mechanism that dramatically increases the flexibility of network has been considered as a deployment strategy rather than the traditional ‘‘bent pipe’’ mechanism [4]. Meanwhile, the delay-sensitive data services are more suitable to transport in the low earth orbit (LEO) satellite networks, which provide inherent advantages in power consumption, pole coverage, latency, and lower cost compared with the geostationary earth orbit (GEO) satellite networks and the medium earth orbit (MEO) satellite networks. Indeed, some commercial companies have scheduled to deploy the IP/MPLS based optical LEO satellite network to supply data services worldwide [9]–[11]. Meanwhile, the IP/MPLS technology has been used for data flows in the terrestrial networks for a long time. The same transport technology facilitates the end-to-end transporting of data flows under unified control in ISTN.

Currently, the data flows in ISTN are considered rigid, and the routing strategy allocates the fixed bandwidth for them in ISTN. However, As a lower convergence of traffic due to the characteristic of the LEO satellite network, the data flow in LEO is fluctuating, which means the data flow requires a basic fixed bandwidth and a shared bandwidth, which we call them elastic bandwidth requirements. Traditionally, if the fixed bandwidth is paired with the elastic data flow, the network utilization will decrease, or the services in the flow will be degraded if the mismatch happens between the fixed allocated bandwidth and elastic bandwidth requirements of the data flow. Compare with terrestrial networks, these negative effects are more severe in ISTN as the resources on the satellite are strictly limited.

On the other hand, the end-to-end delivery of the data flows is better for the QoS and network efficiency [12]. However, most researches on routing mainly focus on the limitation of one network of ISTN, with treating another one as the auxiliary part with rarely considering, and coordinate the two types of networks through the specified gateway between them. Actually, due to the beams between satellite network and terrestrial network are point-to-area coverage, an LEO satellite can link with multiple ground station, and one transport node in terrestrial can have multiple ground station to link with multiple satellites, which means there more optional potential satellite-to-terrestrial links (STLs) exist between satellite networks and terrestrial networks. The end-to-end routing mechanism in ISTN is not as simple as the traditional optimization problem in satellite network or terrestrial networks. It needs to select an appropriate transport path from the set of satellite links, satellite-terrestrial links, and terrestrial links. There are much more challenges to realize an end-to-end routing mechanism in ISTN due to these three types of links have different requirements on dynamics, capacity, latency, fragment, etc. Meanwhile, the LEO satellite network is highly dynamic, which makes the LEO satellite-terrestrial network a hybrid dynamic network together with the fixed terrestrial networks. The hybrid dynamic network makes the routing much more complicated than ever.

In order to realize the end-to-end cooperation on routing between satellite networks and terrestrial networks, the challenges mentioned above are mainly due to the following two new main characteristics between the two types of networks in ISTN:

1) DIFFERENT CHARACTERISTICS OF DIFFERENT LINKS

As the ISTN is hybrid dynamic, the routing mechanism of the satellite networks and terrestrial network have different considerations, and they are relatively isolated with different constraints from each other. In addition to the two networks, the constraints of the STLs are different from the links of the two networks. The traditional routing mechanism only considers the constraints of the satellite links and terrestrial links with a satellite-terrestrial link which is pre-specified. However, the end-to-end routing mechanism should take into account the limitations on the STLs without any specified link.

2) MUCH MORE DYNAMIC POTENTIAL STLs

As mentioned above, there are much more potential STLs between the LEO satellite network and terrestrial networks, which means the data flows can be assigned to any link from the satellite network to local terrestrial network. Meanwhile, the number of the actually used STLs from one satellite or ground station is less than the number of potential links due to the limitation of the number of beams of the satellite or the ground station. As the rotation of the earth and the period of the satellite network, these potential STLs is also dynamic. As the transport path of the traditional routing mechanism is composed of the result of the respective routes of the two networks and the specified STL, the transport path of end-to-end routing mechanism is composed of the links selected from the three sets of different types of links.

Since SDN has achieved the unified control of the network, the QoS attributes and transport paths of SDN-based routing mechanism are also unified managed, which makes it possible to the end-to-end routing under heterogeneous networks. Thus the ISTN can provide the data flow based on the mechanism of request first and then the corresponding path is distributed by the network manager or SDN controller rather than the routing table in traditional IP network, which means the link is based on allocation and is recyclable. In this paper, we focus on the end-to-end routing for elastic data flows in ISTN, with considering the difference of the satellite networks and terrestrial networks, and focus on the new features of ISTN. We first generalize the typical scenario, which consists of a layer-2 based optical LEO network and terrestrial networks, and introduce the necessary modules for unified control and management of the network resources. Then we model the data flows in ISTN and propose a heuristic ant colony based end-to-end fragment-aware path calculation (ACO-EFPC) algorithm to realize unified routing for the elastic data flows. The contributions of this paper are summarized as follows:

1) GENERALIZE THE TYPICAL ISTN SCENARIO AND UNIFIED CONTROL ARCHITECTURE

We generalize a typical network scenario of ISTN, not only contains the LEO satellite network and terrestrial network, but also include the multiple potential STLs between the two networks, the data flow access from any satellite and transport through the two networks and one STLs to the server or data center in terrestrial networks. Meanwhile, for the unified routing mechanism, we generalize an SDN-based unified control architecture and propose the requirements of the necessary modules.

2) MODEL THE ELASTIC DATA FLOW AND UNIFIED ROUTING PROBLEM IN ISTN

We present the elastic data flow model, which defines the properties of the elastic data flows, which are similar with the data flow in the current terrestrial network, include the constraints of the parameters as Committed Reserved Bandwidth (CRB) and Maximum Support Bandwidth (MSB). Based on the scenario and the elastic data flow model, we give out the model of unified end-to-end routing problem in ISTN.

3) THE HEURISTIC ALGORITHM FOR THE UNIFIED ROUTING PROBLEM

We proposed a heuristic path calculation algorithm named ACO-EFPC for the unified routing problem, with considering the characteristics of the latency, wavelength fragment and load balancing to adapt to the integrated network. The algorithm contains two parts which are used for the new incoming data flows and transferring data flows respectively. Then we verify the superiority of ACO-EFPC by the numerical simulation.

The rest of this paper is organized as follows. In Section II, we introduce the related work. The scenario, control architecture, the data flow model, and the unified routing problem is present In Section III. Section IV gives the description and implementation of the ACO-EFPC algorithm. In Section V, numerical simulation to verify the superiority of ACO-EFPC. Finally, Section VI gives out the conclusions.

II. RELATED WORK

The ISTN have been researched for a long time. Whereas it is impossible to operate an integrated network with different systems, protocols, and architectures, software-defined network (SDN) and network function virtualization (NFV) [1] have been introduced in, which offer the opportunity to manage heterogeneous network together [2]. The new paradigm which introduces Satellite Virtual Network Operator (SVNO), and operates the network without concerning the physical infrastructure similar to the network operator in terrestrial networks, has proposed [3]. These research brings much more opportunities for diversity applications runs on the heterogeneous infrastructure. With the multiple satellite-terrestrial links, massive delay-sensitive data flows are supported in various scenarios such as navigation, aircraft, remote area, etc.

In order to realize unified control and management of ISTN, most related researches mainly focus on the architecture design and the network collaboration between the two networks. The research in [12]–[14] present a serial of surveys on the integrated satellite-terrestrial networks, and they introduce the advantage of the integrated network in the application scenario and the SDN-enabled architecture. The satellite networks in HetNet [4] are treated as the core network, together with the core network in terrestrial networks, serve for the edge networks of the terrestrial networks. It gave us a lot of inspiration. However, they just proposed the unified management in the integrated network, rarely concerning on the path calculation problem or routing mechanism in ISTN.

Meanwhile, some schemes used to coordinate the integrated network in different scenarios have proposed to take advantages of both networks. The evolutions that integrate the satellite network with the 5G network [5], [12] Internet of Things (IoT) [6], and Internet of Vehicles (IoV) [7], have proposed the scenarios to coordinate the different terrestrial networks on architecture, management, and protocol. Some of these researchers have proposed the joint routing mechanism. However, they treat the terrestrial networks as the main network and treat the satellite networks as a relay point, without concerning the detail in satellite networks.

For the routing mechanism, we investigate the routing algorithms and other resource allocation algorithms proposed in recent years. Alagoz *et al.* [8] investigate a series of research on routing in satellite networks, and put forward the idea of routing from the view of the integrated satellite network and terrestrial networks. However, it only discusses the possibility of implementing MPLS-based routing in ISTN, without specific research on the routing problem. Li *et al.* [2] present an SDN-based integrated architecture in ISTN and propose a pre-assign algorithm in ISTN. However, the scenario is quite different from considering a simple terrestrial network and a complex satellite network contains Medium Earth Orbit (MEO) and LEO together, and no ISLs between LEO satellites.

Actually, there are two types of research on routing in the integrated network. Some of the studies focus on the routing in the terrestrial network, and they treat the satellite network as a simple relay point of the terrestrial network. A survey on QoS providing of the integrated network [15] emphasized the satellite as a supplement to the ground, and the purpose of optimization mainly focus on the terrestrial network that needs satellite relay [16]. On the other hand, other researches focus on the routing from the perspective of the satellite networks, they just concern on the routing algorithms on satellite networks only, with ignoring the terrestrial network. Wen *et al.* [17], Houtian *et al.* [18], and Dong *et al.* [19] consider the optical LEO satellite network and present the corresponding ant-colony algorithms for routing and wavelength assignment. Liu *et al.* [20], Cheng *et al.* [21], and Yan *et al.* [22] introduce kinds of optimizations on routing problem of the LEO network. This type of research focuses on network efficiency or load balancing. However, they consider the services or data flows which are only in satellite network, with no synergy with the terrestrial network.

As mentioned above, a variety of scenarios have proposed for different applications, and some common characteristics have been proposed in these researches. Thus, a typical scenario arises, in which a network operator can operate the heterogeneous network with a worldwide satellite network, terrestrial networks, and the multiple potential STLs between the two networks, to jointly serve worldwide for data flows.

In this paper, we generalize the scenario of the integrated LEO satellite network and terrestrial networks, which considers both of the LEO satellite network and terrestrial networks, then a routing algorithm for the scenario has been proposed. Although the research results have not completely covered this area yet, they inspire us a lot.

III. SCENARIO, MANAGEMENT ARCHITECTURE AND MODELS

In this chapter, we mainly talk about the generalized scenario of the ISTN, the management architecture for unified control, the models of the elastic data flows, and the model of the routing problem in ISTN. They are described as follows:

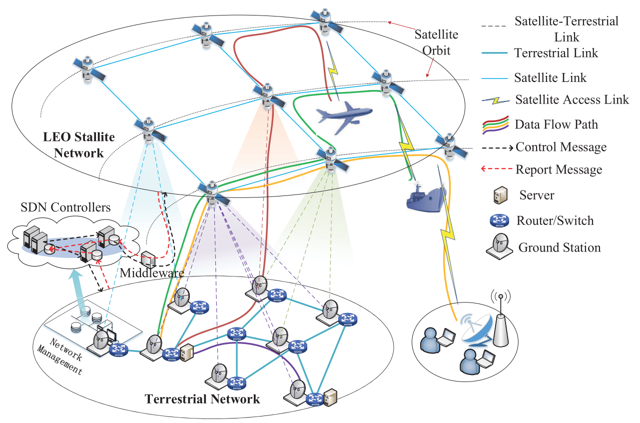

FIGURE 1. The typical scenario of ISTN.

A. THE SCENARIO OF ISTN WITH POTENTIAL STLs

According to the researches talked above, the ISTN mainly contains satellite networks and terrestrial networks. We generalize the scenario of ISTN shown in Fig. 1. The infrastructure of the scenario contains an LEO satellite network and terrestrial networks. As the traditional satellite constellations are stable and widely used compared with the self-reconstructed satellite networking [23], we suppose that the satellite network is formed by an LEO satellite constellation (e.g., Iridium [24], Globestar [25], etc.), whereas the terrestrial network consists of ground stations, switches, and servers. We supposed that the links of satellite network are optical inter-satellite links (ISLs) [26], and the links of the terrestrial network are optical fiber for higher bandwidth, which is widely used. Although some restrictions on STLs such as elevation angle, most constellations guarantee more than one satellite can link with the terrestrial network at any time. Different from the traditional integrated networks, there are multiple time-variant potential STLs because of the satellite can link with multiple ground stations using different beams at the same time. Thus, the data flows transported from the satellite network to the server or data center in terrestrial networks can go through any STL. The network management owns the authority to operate the whole ISTN based on the SDN technology, which makes the unified management of routing and QoS possible. This typical scenario is much more meaningful for lots of applications and scenarios. Note that, the LEO satellite network can be used for data service access, service aggregation and data flow transport with the characteristic of lower latency. Meanwhile, the transport path for the continuous flows changes over time caused by the dynamic STLs and periodic ISLs.

The terrestrial network does not only provide the transport path for the data flows which transport through the satellite network, but also for the data flows which transport only in the terrestrial networks. As the traffic of the data flows transport only in terrestrial networks is much more larger than that in ISTN, and the solution of routing problem is more matured in the pure terrestrial network, we just ignore the data flows which transport only in terrestrial networks, and only consider the data flows transport through the satellite network and terrestrial network simultaneously in this paper.

The centralized management and control architecture of ISTN that have been proposed are based on SDN. The details of the management architecture will be covered in the next section.

As mentioned above, the ISTN can be modeled as a graph $\boldsymbol{G} =< \boldsymbol{V}, \boldsymbol{E} >$, where $\boldsymbol{V}$ is the node set, and $\boldsymbol{E}(t)$ is the link set in ISTN. $e(i,j)$ is the element of $\boldsymbol{E}(t)$ and indicate the status of the link, which equals to 1 if the link between node i and j in $\boldsymbol{V}$ is available, and otherwise is 0. As a detail, the component of $\boldsymbol{G}(t)$ is shown in (1).

$$\boldsymbol{G}(t) = \boldsymbol{G}_s(t) \cup \boldsymbol{G}_t \cup \boldsymbol{E}_{st}(t). \tag{1}$$

where $\boldsymbol{G}_s(t) =< \boldsymbol{V}_s, \boldsymbol{E}_s(t) >$ and $\boldsymbol{G}_t =< \boldsymbol{V}_t, \boldsymbol{E}_t >$ are the network graphs of satellite network and terrestrial network respectively. Meanwhile, $\boldsymbol{E}_{st}(t)$ is the links set between $\boldsymbol{G}_s(t)$ and $\boldsymbol{G}_t$. Both $\boldsymbol{E}_s(t)$ and $\boldsymbol{E}_{st}(t)$ are time-variant, which own the element $e_s(i,j)$ and $e_{st}(i,j)$ to indicate the link availability between i and j, whereas $e_t(i,j)$ is the element of $\boldsymbol{E}_t$ for indicating the link availability from node i to node j which is the element of $\boldsymbol{V}_s$. The link availability indicator variable equals to 1 when the corresponding link is available, and otherwise is 0.

Based on the characteristics of data flows and the present satellite networks, we assume that the ISLs in ISTN are based on IP/MPLS over wavelength division multiplex (WDM) with the optical link. The capacity of the ISL optical link is indicated by (2):

$$\begin{cases} B_s = \sum_{i=1}^{W} \lambda_i^s = W\lambda^s \\ \lambda_1^s = \lambda_2^s = \ldots = \lambda_w^s = \lambda^s \end{cases} \tag{2}$$

Where B_s is the bandwidth of each ISL, and λ_i^s is the bandwidth provided by i-th wavelength which equals λ^s, and W is the number of wavelengths in each ISL. Meanwhile, the STLs in ISTN are based on microwave-band (e.g. Ka, Ku, etc.), which multi-carriers or multi-beams is supported by the satellite. Then the capacity of STL can be described by (3):

$$\begin{cases} B_{st} = \sum_{i=1}^{U} \lambda_i^{st} = U\lambda^{st} \\ \lambda_1^{st} = \lambda_2^{st} = \ldots = \lambda_U^{st} = \lambda^{st} \end{cases} \tag{3}$$

Where B_{st} is the bandwidth of each STL, λ_i^{st} is the bandwidth of the i-th beams which equals λ^{st}, and U is the number of beams in each STL.

B. SDN-BASED MANAGEMENT ARCHITECTURE

In order to better fit our proposed scenario talked above as well as the elastic data flow model and the routing algorithm mentioned later, we describe a general control architecture, and some of the necessary functions or modules shown in Fig. 2.

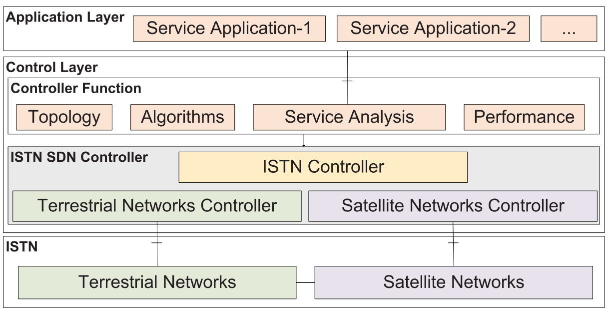

FIGURE 2. The necessary management and control architecture of SDN-based ISTN.

As ISTN contains terrestrial networks and satellite networks, the SDN controller, which runs in the network management server, realizes unified control of the two heterogeneous networks through the southbound interface by the hierarchical management architecture. A high-level controller coordinates the two low-level controllers which used to control satellite network and terrestrial networks respectively.

The controller function unit is attached to the controller and realizes the centralized routing mechanism for different flows from the application layer. Whereas the topology function is used to manage the dynamic topology and slice it into different topologies with continuous time slice, the service analysis module is used to standardize the flow request and return the final path to the different flow requests from different service applications. The algorithms module used to execute the key algorithm for allocating bandwidth, whereas the performance module used to examine the status and the numerical analysis of ISTN network. Meanwhile, the southbound interface of the terrestrial controller can use the standard protocol (e.g. OpenFlow, etc.), while that of the satellite network controller can be realized through middleware such as GEO satellite [2], ground station and so on. The controllers collect the information of ISTN using report message and send instructions to the networks with control message. The two types of messages are also shown in Fig. 1. Both of them can be nested in the standard SDN protocol. The SDN protocol is either in-band or out-of-band, depending on the actual deployment. With the coordination of these basic functions and the control architecture, the end-to-end routing mechanism can be realized.

Since the original SDN protocol (OpenFlow) lacks QoS-related messages, there are several feasible solutions to configure the QOS of the network devices. While the switches in the network are SDN-enabled by Open vSwitch (OvS), the configuration of the switch can be realized by the OvS command [29] remotely. Meanwhile, the MPLS OpenFlow protocol extensions [30] with complete QoS solution has become a formal standard. The SDN-based research supporting QoS functions has also been studying a lot [31].

Note that, our algorithm for routing is based on the unified management of ISTN, and is independent on the way to achieve unified management. Although we describe the architecture of unified management in ISTN, the actual architecture can include but is not limited to it.

C. ELASTIC DATA FLOW MODEL

Different from the traditional service in ISTN, most data flows are elastic and continuous, which means the traffic of the service in the data flows is time-variant and long-term. We With the routing mechanism of request first we talked above, we model the flow request set as (4):

$$\begin{cases} F(t) = \{f_i(t) | i \in [1, K]\} \\ f_i(t) = [T_b^i, T_e^i, S_i(t), D_i, B_c^i, B_m^i] \\ P_i(t) = P_i^s(t) \cup P_i^{st}(t) \cup P_i^t(t) \end{cases} \tag{4}$$

Where $F(t)$ is the set of the data flow requests, $f_i(t)$ is the i-th data flow request in $F(t)$, and K is the number of the requests. The properties of $f_i(t)$, where T_b^i and T_e^i are the begin time and end time of the data flow, is also defined by (4). $S_i(t)$ and D_i are the two endpoints of the i-th data flow, in which $S_i(t)$ indicates the endpoint which connects to the satellite network and D_i indicates the endpoint in the terrestrial network. The $S_i(t)$ is time-variant and D_i is static because of the inherent characteristics of the dynamic satellite network and the static terrestrial networks. B_c^i and B_m^i are the constraints as CRB and MSB. $P_i(t)$ represents the set of the transport path of $f_i(t)$. $P_i^s(t)$, $P_i^{st}(t)$ and $P_i^t(t)$ represent the different segments as satellite network, STLs and terrestrial network of $P_i(t)$.

For each data flow, we allocate exclusive bandwidth and shared bandwidth on the corresponding links, in which the shared bandwidth resource can be used by all of the data flows on the link, but they are limited by B_c^i and B_m^i. In detail, B_c^i indicate the minimize bandwidth the network should provide for $f_i(t)$, and B_m^i of $f_i(t)$ indicate the peak used bandwidth of the i-th data flow. These two parameters are given by the data flow request. They show the elasticity of the network flow, that the links with the unallocated bandwidth should greater than or equals to B_c^i, and the shared bandwidth in the link keeps the bandwidth of i-th data flows in the link can reach to B_m^i if they used to transport $f_i(t)$. In other words, $f_i(t)$ owns its exclusive bandwidth as B_c^i, and can use the extra bandwidth from the shared bandwidth of the links, but the total used bandwidth of $f_i(t)$ cannot exceed B_m^i. Thus, the relation of the link bandwidth and $F(t)$ is as (5) shows.

$$\begin{cases} q_{ij}(t) \le b_{ij} \\ \upsilon_{ij}(t) = \sum\limits_{\substack{f_x \in F(t) \\ p_{ij}(x,t) \in P_x(t)}} p_{ij}(x,t) B_c^x \\ \tau_{ij}(t) = \max\limits_{\substack{f_x \in F(t) \\ p_{ij}(x,t) \in P_x(t)}} \left(p_{ij}(x,t)\left(B_m^x - B_c^x\right)\right) \\ q_{ij}(t) = \upsilon_{ij}(t) + \tau_{ij}(t) \end{cases} \tag{5}$$

Where b_{ij} represents the bandwidth of $e(i,j)$. $p_{ij}(x,t)$ is the element of $P_x(t)$, and indicates that the link $e(i,j)$ is distributed for the i-th flow at time t if the value is 1. $\upsilon_{ij}(t)$ and $\tau_{ij}(t)$ means the total CRB and total elastic bandwidth (means the shared bandwidth on the link, which equals the value of MSB minus the value of CRB) of the data flows in link $e(i,j)$ respectively. $q_{ij}(t)$ represents the demand bandwidth of $e(i,j)$ at time t. For a better understanding, Fig. 3 shows the relationship of these different parameters and the influence on the amount of allocated bandwidth in the network.

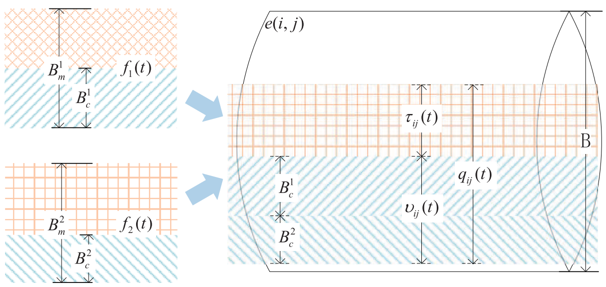

FIGURE 3. The bandwidth parameters of elastic data flow model.

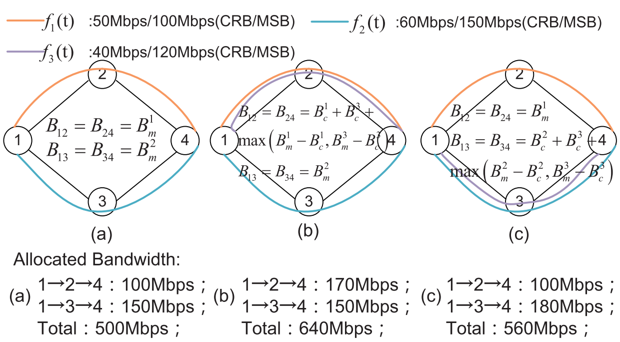

FIGURE 4. The example of allocated bandwidth.

Note that, the amount of the allocated bandwidth will be different due to the different selected paths, and this is mainly decided by elastic bandwidth $\tau_{ij}(t)$. As an example, the Fig. 4(a) shows a status of $f_1(t)$ and $f_2(t)$ transport from 1 to 4, with two different paths as 1→2→4 and 1→3→4, and the total amount of the allocated bandwidth in Fig. 4(a) is 500Mbps. When $f_3(t)$ arrives from 1 to 4, there are two paths can be selected, if $f_3(t)$ is assigned in 1→2→4, which shown in Fig. 4(b), the total amount of allocated bandwidth is up to 640Mbps, and this value will be 560M if $f_3(t)$ is assigned in 1→3→4 shown in Fig. 4(c). The path 1→3→4 for $f_3(t)$ is better than the path 1→2→4 in the utilization of the network as the path 1→3→4 use the less amount of allocated bandwidth and transport the same flows as 1→2→4 in the network.

As mentioned, we assume that the traffic of data flows transport through the satellite network is much smaller than that transport in the terrestrial network only. Thereby, we suppose that the terrestrial network can satisfy the traffic of data flows in ISTN without any constraint besides latency.

D. THE UNIFIED ROUTING MODEL

As the trajectory of the satellite in the constellation is predictable with the constantly changing topology, it is adaptive to using virtual topology to realize the routing mechanism. Thus, the topology of ISTN can be discretized into lots of temporary topologies expressed in (6):

$$R(t) = [R(t_1), R(t_2), \ldots, R(t_i), \ldots, R(t_T)] \tag{6}$$

Where T is the total number of the virtual topologies of the $R(t)$, $R(t_i)$ is the i-th topologies in $R(t)$. Then, we give a related definition of ISTN.

Definition 1: The minimum bandwidth of STLs for data flows is $\Delta_k(t_i, t_j)$, $i \le k \le j$, between time t_i and t_j. In which

$$\begin{cases} \Delta_k(t_i, t_j) = B_{st} C_{\min}(t_i, t_j) \\ C_{\min}(t_i, t_j) = \min\limits_{t_x \in [t_i, t_j]} \sum\limits_{e(i,j) \in E_{st}(t_x)} e(i,j) \\ = \sum\limits_{e(i,j) \in E_{st}(t_k)} e(i,j) \end{cases} \tag{7}$$

where B_{st} represents the bandwidth of the STL. $C_{\min}(t_i, t_j)$ is the minimum number of available STLs between time t_i and t_j, $\Delta_k(t_i, t_j)$ indicates the minimum bandwidth of the transport capacity between the two networks, and the corresponding time is t_k. The definition gives out the upper limit of the capacity of ISTN.

We proposed the unified routing algorithm with considering the following factor:

1) LATENCY OF THE FLOWS IN ISTN

As the dynamic of the topology, the ISTN cannot guarantee the stable latency, so we suppose that the ISTN should provide the best effort on latency. Meanwhile, we suppose the latency is fixed and equals the maximum value of the interval in each slicing time, to ensure that the required latency of the data flow can be guaranteed.

2) WAVELENGTH FRAGMENTATION IN LEO AND STLs

Taking into account the limitations of the satellite we mentioned above. The fewer wavelength fragmentation means higher efficiency. Meanwhile, it can extend the lifetime of the onboard devices of the satellite. In the dynamic topology, this should be an important issue in satellite network and STLs.

3) TOTAL ALLOCATED BANDWIDTH

With the elastic data flow model we talked in subsection *C* of this chapter, different selected path occupy different total allocated bandwidth in the network. So we hope that the network can provide the corresponding bandwidth for flows with the least total allocated bandwidth.

4) LOAD BALANCING OF SATELLITE NETWORKS

As the difficulty of maintaining satellite networks, the load balancing mechanism can avoid part of the devices on the satellite from aging in advance. A better load balancing mechanism is conducive to the lifetime of the devices in the satellite networks.

Aimed above, our object is to minimize the evaluation function as (8):

$$\Theta(t) = \omega_1 l(t) + \omega_2 F_r(t) + \omega_3 \sigma^s(t) + \omega_4 \mu^s(t) + \omega_5 \eta(t) \tag{8}$$

In which:

$$\begin{cases} l(t) = 1 - e^{\left(-\frac{\sum_{f_i \in F(t)} L_i^{ant}(t)}{\sum_{f_i \in F(t)} L_i^{\min}(t)}\right)} \\ F_r(t) = mF_r^s(t) + nF_r^{st}(t) \\ F_r^s(t) = \dfrac{\sum_{\substack{f_x(t)\in F(t) \\ p_{ij}^s(x,t)\in P_x^s(t)}} (\lambda_s - \mathrm{mod}(q_{ij}^s(t), \lambda_s))}{\sum_{e(i,j)=1} \lambda_s} \\ F_r^{st}(t) = \dfrac{\sum_{\substack{f_x(t)\in F(t) \\ p_{ij}^{st}(x,t)\in P_x^{st}(t)}} (\lambda_{st} - \mathrm{mod}(q_{ij}^{st}(t), \lambda_{st}))}{\sum_{e(i,j)=1} \lambda_{st}} \\ \sigma^s(t) = \sigma\left[q_{ij}^s(t)\right], f_x \in F(t), p_{ij}^s(x,t) \in P_x^s(t) \\ \mu^s(t) = \dfrac{\max\limits_{p_{ij}^s(x,t)\in P_x^s(t)} (p_{ij}^s(x,t)) - \min\limits_{p_{ij}^s(x,t)\in P_x^s(t)} (p_{ij}^s(x,t))}{\lambda_s} \\ \eta(t) = \dfrac{\sum_{p_{ij}^s(x,t)\in P_x^s(t)} p_{ij}^s(x,t)(\phi_{ij}(t) - \phi_{ij}(t-1))}{\sum_{p_{ij}^s(x,t)\in P_x^s(t)} p_{ij}^s(x,t) B_m^x} \end{cases} \tag{9}$$

$$\text{Subject to: } \begin{cases} \forall q_{ij}^s(t) \le B_s \\ \forall q_{ij}^{st}(t) \le B_{st} \\ \sum q_{ij}^{st}(t) \le \Delta_k(t_i, t_j) \quad , e(i,j) \in \boldsymbol{G}(t) \\ \sum_{i \in \boldsymbol{G}_t(t)} e(i,j) \le U \end{cases} \tag{10}$$

In (8), ω_1, ω_2, ω_3, ω_4, and ω_5 represent the weight of the parameters, in which $l(t)$ is the latency factor, $F_r(t)$ is the fragment factor, $\sigma^s(t)$ is the variance factor, $\mu^s(t)$ is the norm factor, and $\eta(t)$ is the MSB overflow factor. Equation (9) shows the calculation formula of the parameters of the evaluation function. Where $P_x^s(t)$ and $P_x^{st}(t)$ are the set of transport path for the x-th flow in $G_s(t)$ and $E_{st}(t)$. $l(t)$ is the latency factor, in which $L_i^{ant}(t)$ is the latency of path for $f_i(t)$, whereas $L_i^{\min}(t)$ is the latency of the path calculated by Dijkstra [27] algorithm with only considering the latency. The $L_i^{ant}(t)$ and $L_i^{\min}(t)$ consist of the time-variant propagation delay and static processing delay. $q_{ij}^{st}(t)$ is the total used bandwidth in one STL, and $q_{ij}^s(t)$ is that in one ISL, which all of them cannot exceed the bandwidth of the corresponding links. As the satellite network is based on WDM, $F_r(t)$ represents the wavelength fragment ratio in satellite network and STL, which respectively expressed by $F_r^s(t)$ and $F_r^st(t)$ with the weight m and n. The lower $F_r(t)$ is, the higher the network utilization and energy efficiency. $\sigma^s(t)$ represents the variance of $q_{ij}^s(t)$, and $\mu^s(t)$ means the difference between the maximum allocated bandwidth and the minimum allocated bandwidth, which used to achieve load balancing, to avoid lifetime problems caused by the overuse of each link. $\eta(t)$ indicates the ratio of the actual increment and the maximum increment of the shared bandwidth whereas the service is served, in which $\phi_{ij}(t)$ means the elastic bandwidth of $e(i,j)$ at time slice t. It used to guide the flows transported in a more suited link to avoid wasting of the bandwidth. Note that, the $\sigma^s(t)$ and $\mu^s(t)$ are both used for load balancing. When a data flow need to be routed, the $\sigma^s(t)$ guides the route to make the loads of the satellite network close to the average value. Once the network takes into account the links with the largest loads or never used, $\mu^s(t)$ will further increase the difficulty of choose these links. In other words, $\sigma^s(t)$ focus on the overall effect and $\mu^s(t)$ used to avoid the worst effect in the links.

The constraints shown in (10) represent the limitation of bandwidth in the links, which means the capacity of the ISTN should be enough to transport data flows. U limits the used STL from the potential STLs, and it is converted from the limitation of the number of used STLs of potential STLs at the same time.

What needs to be emphasized is, with the different characteristics of the three links mentioned above, the $l(t)$ is calculated using the whole path in ISTN, whereas $F_r(t)$, $\sigma^s(t)$ and $\mu^s(t)$ only using the path in $G_s(t)$ and $E_{st}(t)$. This is because of the abundant flows transport only in the terrestrial network makes these factors for terrestrial inaccurate and invalid.

As $\Theta(t)$ contains several independent parameters, it is considered as an NP-complete problem [28] to calculate the transport path. As the link status can be predicted for the law of satellite movement, and the routing just require the path calculation result before the corresponding slicing time is up. It is a feasible solution to use a heuristic algorithm to solve this problem. Meanwhile, the ant colony algorithm as a matured heuristic algorithm to solve NP problem has been applied in many industries and fields [12]–[14]. As a result, we propose the ant colony based heuristic algorithm named the ACO-EFPC, to solve our unified routing problem. The detail description of ACO-EFPC is in the following chapter.

IV. THE MECHANISM OF ACO-EFPC

Based on the above analysis, we use an ant colony-based heuristic algorithm to search for the optimum path in ISTN. As a result, ACO-EFPC can adapt to the dynamic changes of the ISTN by the effect adaptive learning of pheromone. Due to the routing mechanism and the scenario we supposed, the proposed algorithm can be competent for path calculation of the routing mechanism. The detailed design and mechanism of ACO-EFPC are as follows:

A. TRANSITION PROBABILITY FUNCTION

In ACO-EFPC, The related notation used in the ant colony algorithm is as follows:

$\tau_{ij}^{k}(t)$: The pheromone value for ant k on the path $e(i,j)$ at moment t.

$Tabu_k(t)$: The tabu list, which registers the nodes that ant k has visited. Besides, once the ant has visited the node in the terrestrial network, all the nodes in satellite network should be registered in the tabu list.

$d_{ij}^{k}(t)$: The latency factor of link $e(i,j)$ for the k-th ant, which equals the ratio of the maximum value of the latency of $e(i,m)$,$m \in G(t)$ and the latency of $e(i,j)$ for the k-th ant..

$\eta_{ij}^{k}(t)$: The congestion factor of the k-th ant on $e_s(i,j)$, which equals the ratio of the remaining bandwidth after supporting the service and total bandwidth of $e_s(i,j)$. It indicates the congestion status of the bandwidth fragment for different path the k-th ant select.

$\upsilon_{ij}^{k}(t)$: The current fragment ratio of bandwidth for ant k, which equals the ratio of the used bandwidth in the not completely used wavelength and the bandwidth of wavelength. It indicates the amount of the bandwidth fragment for the path $e_(i,j)$ which the k-th ant select.

$\rho_{ij}^{k}(t)$: The MSB overuse factor of $e(i,j)$, which indicates the degree of matching between the current elastic bandwidth of $e(i,j)$ and the elastic bandwidth of the current flow, as shown in (11)

$$\rho_{ij}^{k}(t) = \begin{cases} e^{-\frac{(B_m^x - B_c^x)-\phi_{ij}(t)}{B_m^x - B_c^x}}, & B_m^x - B_c^x > \phi_{ij}(t), x \in [1,K] \\ 1, & else \end{cases} \tag{11}$$

$$P_{ij}^{k} = \begin{cases} \dfrac{[\tau_{ij}^{k}(t)]^{\alpha}\cdot[d_{ij}^{k}(t)]^{\beta}\cdot[\eta_{ij}^{k}(t)]^{\gamma}\cdot[\upsilon_{ij}^{k}(t)]^{\varsigma}\cdot[\rho_{ij}^{k}(t)]^{\xi}}{\sum\limits_{m\in\Upsilon_k}[\tau_{im}^{k}(t)]^{\alpha}\cdot[d_{im}^{k}(t)]^{\beta}\cdot[\eta_{im}^{k}(t)]^{\gamma}\cdot[\upsilon_{im}^{k}(t)]^{\varsigma}\cdot[\rho_{im}^{k}(t)]^{\xi}}, \\ \qquad i \notin V_s, \quad j \in \Upsilon_k,\, j \notin Tabu_k \\ 0, \qquad else \end{cases} \tag{12}$$

In which

$$(\alpha,\beta,\gamma,\varsigma,\xi) = \begin{cases} (\alpha_1,\beta_1,\gamma_1,\varsigma_1,\xi_1), & i \in V_s, j \in V_s \\ (\alpha_2,\beta_2,\gamma_2,\varsigma_2,\xi_2), & i \in V_s, j \in V_t \\ (\alpha_3,\beta_3,\gamma_3,\varsigma_3,\xi_3), & i \in V_t, j \in V_t \end{cases} \tag{13}$$

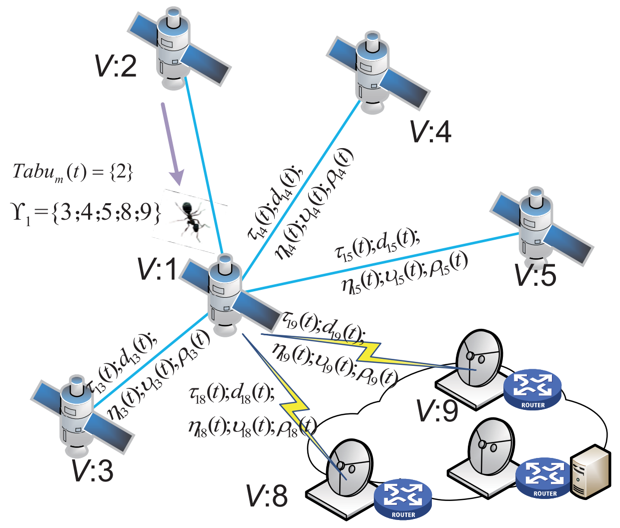

FIGURE 5. The diagram of the parameter and heuristic function.

Equation (12) shows the transition probability function of ACO-EFPC. Υ_k is the node set which can be selected for ant k at node i, and α is the information heuristic factor and represents the significance of the tracks made by the former ants, β, γ, ς and ξ are the expectation heuristic factors, which represent the significance of $d_{ij}^{k}(t)$, $\eta_{ij}^{k}(t)$, $\upsilon_{ij}^{k}(t)$ and $\rho_{ij}^{k}(t)$ respectively. P_{ij}^{k} represents the probability to choose $e(i,j)$ for the next path of ant k. Since there are three types of links in ISTN, these expectation heuristic factors have three values respectively corresponding to different types of links. As an example, the parameters and heuristic functions in ISTN shows in Fig. 5, there is an ant at V:1 from V:2, at this moment, then the V:2 is in the tabu list of this ant, and the Υ_k updates for the possible path to choose. Then the current $d_{ij}^{k}(t)$, $\eta_{ij}^{k}(t)$, $\upsilon_{ij}^{k}(t)$ and $\rho_{ij}^{k}(t)$ are sent to the k-th ant, the ant k determine the next path through the transition probability calculated by the corresponding heuristic factors.

1) PHEROMONE STRATEGY

Once all of the ants finish the path, the pheromone must be updated for the latter ants using this information. As the typical approach, the ants which find the destination enhance the pheromone they passed, whereas the pheromone of other paths is partly evaporation. If the current value of pheromone is $\tau_{ij}(t)$, then the pheromone of the next generation can be calculated as (14):

$$\begin{cases} \tau_{ij}(t') = (1-\rho)\tau_{ij}(t) + \Delta\tau_{ij}(t'), \rho \in [0,1] \\ \Delta\tau_{ij}(t') = \sum\limits_{k=1}^{m} \tau_{ij}^{k}(t) \\ \tau_{ij}^{k}(t) = 1 - \dfrac{\Theta_k(t)}{\max\limits_{i\in[1,m]} \Theta_i(t)} \end{cases} \tag{14}$$

Where m represents the number of ants, ρ is the evaporation coefficient of the pheromone, which decides the influence of the history information. $\Theta_k(t)$ represents the cost $\Theta(t)$ of ant k. As $\Theta_k(t)$ smaller is better, $\tau_{ij}^k(t)$ of better ant k will turn bigger. Finally, the corresponding $\tau_{ij}(t')$ will be bigger than others. As a result, the latter ants get more probability to this path.

2) THE ACO-EFPC

Considering the difference between the satellite network and terrestrial networks we mentioned in section II, we make two assumptions:

First, as the transport capacity of the terrestrial network is much larger than that of the satellite network, we assume that the data flows in ISTN have negligible influence on terrestrial network segment of routes, what means we only consider the latency in the terrestrial network.

Second, as the links between the satellite network and terrestrial network are limited, and the capacity of ISTN is limited by the capacity of the available STLs, we assume that one data flow in ISTN only transports through one STL and transport through the STL once, what is pretty obvious.

Algorithm 1 ACO-EFPC Algorithm at Time t_a for New Incoming Data Flow

Input:
The set of flow requests, $F(t_a)$;
The topology of the ISTN, $G(t_a)$;
Output:
The set of the transport path of the data flows, $\Omega(t_a)$;
1: **for** each $f_i(t_a) \in F(t_a)$ **do**
2: **if** (10) is satified **then**
3: Initial ant nums,$\phi_{ij}^k(t_a)$ and *Gens*
4: **for all** ant k in ants **do**
5: Caculate $\eta_{ij}^k(t_a)$, $\upsilon_{ij}^k(t_a)$, $d_{ij}^k(t_a)$ and $Tabu_k(t_a)$
6: **for** each node $j \in \Upsilon_k$ **do**
7: **if** $j \in G_s(t_a)$ **then**
8: Run Dijkstra algorithm for shortest path of j and endpoint of $f_i(t_a)$ on the terrestrial network
9: The k-th ant choose the path through the result of Dijkstra algorithm.
10: **else**
11: Calculate P_{ij}^k using (11)
12: Select the path for k-th ant by P_{ij}^k
13: **end if**
14: Update pheromone by (12)
15: Save the best path in $R(t_a)$ by (8)
16: **end for**
17: **end for**
18: **end if**
19: **end for**
20: return $\Omega(t_a)$

ACO-EFPC contains two sub-algorithm which are used for the new incoming data flows and forwarding data flows respectively. The procedure of the ACO-EFPC for new incoming data flows at time t_a is shown in Algorithm 1. For efficient calculations, the ACO-EFPC has joint the ant colony algorithm and Dijkstra algorithm together. The main idea is once one ant goes into the terrestrial network, the ant will stop to search and finish the remaining path determined by Dijkstra algorithm, with the source node and the destination node is the current position of ant and the destination of the data service in terrestrial network. This process is shown in *Step*7 to *Step*9. However, the whole path will be evaluated by the cost function of our model, and then the pheromone will be updated in $G_s(t)$ and $E_{st}(t)$. Meanwhile, with considering the characteristics of dynamic topology, the *Algorithm* 2 will be triggered when the source of the service changed, or the origination route is a failure.

Algorithm 2 ACO-EFPC Algorithm at Time t for the Flow Being Forward

Input:
The set of the path of forwarding data flow at time t_{a-1}, $R(t_{a-1})$;
The topology of the ISTN at time t_{a-1}, $G(t_{a-1})$;
The topology of the ISTN at time t_a, $G(t_a)$;
Output:
The set of the transport path of the data flows at time t_a, $\Omega(t_a)$;
1: **for** each $P_i(t_{a-1})$ **do**
2: **for** each $e(i,j) \in G(t_{a-1})$ in $P_i(t_{a-1})$ **do**
3: **if** $e(i,j) \in G(t_a)$ equals to 0 **then**
4: Romove data flow $f_i(t_{a-1}) \in \boldsymbol{F}_i(t_{a-1})$
5: **for** each $e(i,j) \in G(t_{a-1})$ in $P_i(t_{a-1})$ **do**
6: Calculate $\upsilon_{ij}(t_{a-1})$, $\tau_{ij}(t_{a-1})$ and $q_{ij}(t_{a-1})$ using (5)
7: **end for**
8: Set $f_i(t_a) = f_i(t_{a-1})$
9: Calculate $P_i(t_a)$ through *Algorithm* 1 using input: $f_i(t_a)$ and $\mathbf{G}(t_a)$
10: **else**
11: $P_i(t_a) = P_i(t_{a-1})$
12: **end if**
13: **end for**
14: **end for**
15: return $\Omega(t_a)$

Algorithm 2 shows the ACO-EFPC for the forwarding flows used to deal with the forwarding flows in ISTN. While one flow needs to calculate a new path, the algorithm will delete the flow from the set of the data flow temporarily, and recalculate the relevant parameters of the network, as shown in *Step*4 to *Step*7. Then recalculate the new path for the data flow using *Algorithm* 1 which shows in *Step*8 and *Step*9, to make sure the path update. Note that, all of the procedure are runs on the server for the path calculation of routing. They do not effect on the actual transport path in ISTN until the routing strategy is delivered to the physical network by the unified management architecture we talked in section III.

V. NUMERICAL SIMULATION AND ANALYSIS

In this chapter, we first introduce the detail simulation scenario, based on this, we analyze the numerical simulation results to prove the superiority of the performance of ACO-EFPC. To better demonstrate the superiority of the algorithms, we divided the simulation into serval parts.

A. SIMULATION SETUP

In our scenario, an LEO satellite network and a terrestrial network have been simulated.

TABLE 1. The parameters of the simulation constellation of LEO.

Parameters	Value
Configuration	π
Inclination,(θ)	86.4°
Number of Satellites,(H)	66
Number of Orbits, (M)	6
Number of Satellite/Orbit, (N)	11
Phasing Factor, (PF)	0
Altitude, (Unit:km)	780
ISL	2 intra/2 inter
ISL state	Non-permanent
Elevation Angle	8.2°

As a typical polar (π) constellation, the Iridium-like constellation is chosen as the constellation of LEO satellite network in the simulation. Table. 1 summarizes the parameters of the constellation we used in this paper. With the characteristics of non-permanent ISLs and no cross-seam ISLs, it is enough complexity to represent the LEO satellite network. As a typical terrestrial network, the parameters of the backbone network of China Education and Research Network (CERNET) are chosen for simulating the terrestrial network. Fig. 6 shows the simulation topology in one time interval.

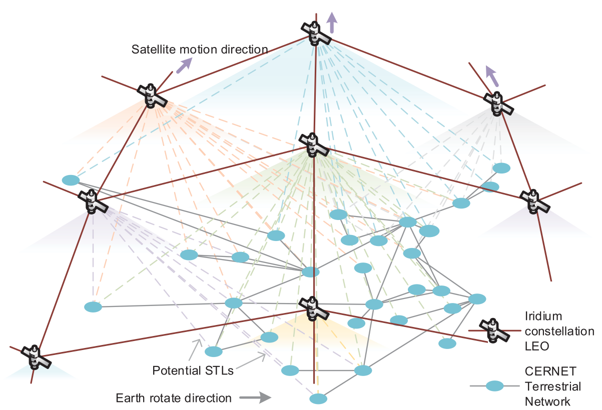

FIGURE 6. The simulation topology in a time interval.

For simulating the STLs, we suppose that the main node in CERNET have deployment the ground station so that the satellite can connect to any ground station once satisfy the connection restrictions. Meanwhile, as to be the upcoming LEO satellite network, we choose the parameters of LeoSat [10] as a reference of the ISLs and STLs. Table. 2 shows the parameters of the ISLs and STLs we used in the simulation.

TABLE 2. The parameters of ISLs and STLs.

Parameters	Value	Parameters	Value
λ^s	800Mbps	λ^{st}	960Mbps
W	3	U	3
B_s	2400Mbps	B_{st}	2880Mbps

As the speed of the LEO satellite is much faster than ground objects, we suppose that the position of the access point of service is static, which has no influence on the result of the simulation. But the access satellite of each data flow is time-variant and iteratively.

In order to generate the time-variant data of the access satellite, ISLs, and STLs, we pre-calculate the satellite model and numerical analysis to get the status and the latency of the ISLs and STLs, and the $S_i(t)$ of $f_i(t)$. The value of all these parameters in each time slice can be calculated in advance. Considering the latency of services contains transport time and processing time, we assume that the processing time of the node in $\boldsymbol{G}(t)$ is static and set it to 100 μs.

As the constantly changing of the states of the ISLs and STLs, we assume that the latency of the $e(i, j)$ in one time slice is static, which is a common method to approximate on routing problem of the satellite network.

Due to the scenario of our model, The data flow may transport from any place in the world, so we generate $F(t)$ randomly.

Based on the data generated by the software and the parameters we set, the simulation of ACO-EFPC is simulated by using MATLAB.

B. SCENARIO CAPACITY

Taking into account the different scenarios as well as the characteristics of data flows in ISTN, we first compare the capacity of the traditional scenarios with one specified STL and the integrated ISTN scenarios with multiple potential STLs.

We create 300 elastic data flow requests, each data flow owns one endpoint which is randomly selected in the worldwide and another endpoint which is in the terrestrial network. Meanwhile, we supposed both the MSB and CRB of these data flows are subject to average distribution as the average value are 150Mbps and 75Mbps, which are the common parameters.

We put these flow requests into 20 consecutive time slices to examine the capacity of different scenarios by using the Dijkstra algorithm. The result is illustrated in Fig. 7. Due to the different topologies with different time intervals, the capacity of both scenarios changes over time. Meanwhile, the capacity of ISTN scenario has significantly improved with the multiple ISLs and multiple gateways compare with the traditional scenario. In our simulation results, the number of accepted services in ISTN scenario is 24.48% to 80.15% higher than that of traditional scenarios.

C. ACO-EFPC PERFORMANCE

In order to check the advantages of ACO-EFPC, we choose one of the compared algorithms named SADR [22], and another parallel algorithm which is Dijkstra algorithm with considering the same factors (latency, wavelength fragmentation, elastic bandwidth, allocated bandwidth) as ACO-EFPC to form a new metric. As the it is hard to precalculate the variance and norm of each transport path for Dijkstra algorithm, we use value of allocated bandwidth as the load balancing factor of Dijkstra algorithm. These two algorithms for comparison are also extended in the ISTN scenario, with the same weight of factors. In the simulation, we select 150 flows randomly from the services set used above. To reduce the amount of calculation, the transport path of the flows will not interrupt until the original path is unavailable. We track the changes of indicators in 20 consecutive virtual topologies.

What is more, it is the situation that only reflect the performance of the fixed load as 150 services, and cannot fully explain the superiority of the algorithm under various utilization of the network. Due to this problem, we test the results of different algorithms by setting different numbers of services from 30 to 180 with an interval of 30, and the services are selected randomly from the services set talked above. Note that, we set the upper limit to 180 because it has reached the capacity of the ISTN.

Now, we analyze the performance of the algorithms on latency, fragment, capacity and load balancing as follows:

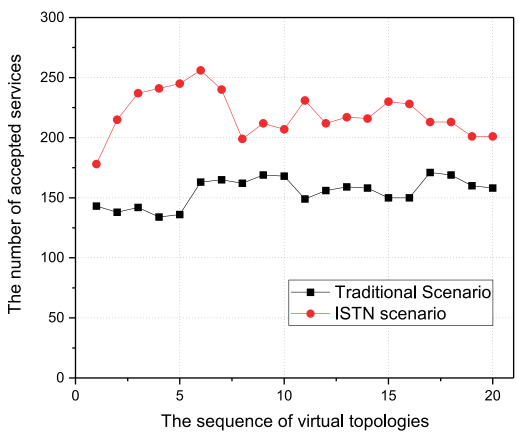

FIGURE 7. The number of accepted services.

1) AVERAGE LATENCY OF ACCEPTED SERVICES

As partly different exists between the two time-continuous virtual topologies, every time the data flows enter a new time slice, only part of the path of the data flows are recomputed. Hence, the latency as well as other indicators fluctuate greatly between the virtual topologies. The performance of the algorithms on latency is shown in Fig. 8 while the number of flows is 150. The average latency of the data flows using the SADR and ACO-EFPC is almost the same, whereas the Dijkstra algorithm has a little disadvantage than the other two algorithms. Note that, as the Dijkstra algorithm uses the new metric with different factors such as wavelength fragment, load balancing, etc., the average latency of data flows using the Dijkstra algorithm is not the lowest. Meanwhile, the Dijkstra algorithm has shown a higher latency at the beginning several time intervals. This is because there are not any data flow in ISTN and other factors in the new metric effect the results of the Dijkstra algorithm at the beginning. As the time interval iteratively, the latency of the Dijkstra algorithm drop slowly and stay at the same level as SADR and ACO-EFPC algorithms, which shows a slow convergence performance of the Dijkstra algorithms in ISTN. In general, the average latency of data flows using ACO-EFPC is 7.52% lower than that using Dijkstra algorithm. Note that, the latency of Dijkstra algorithm is not lower than that of ACO-EFPC after convergence (3.42% higher than ACO-EFPC after the 10-th virtual topology).

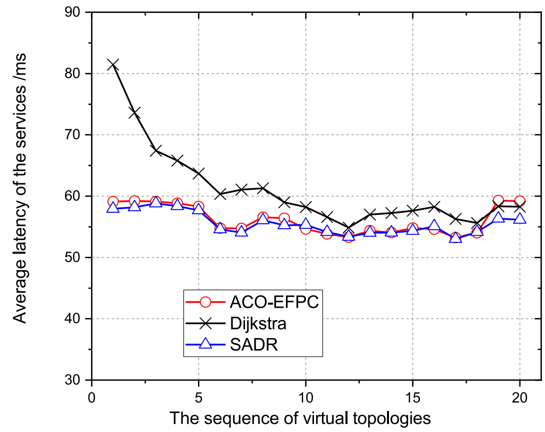

FIGURE 8. The average latency of 150 data flow requests.

The average latencies of different numbers of flow requests in 20 virtual topologies using different algorithms are shown in Fig. 9, When the number of flow requests is small (lower load), the data flows have more free paths to select, as the load increases, the data flows have fewer paths to choose from because of the block of part of the links in ISTN. Thus, as the number of flow requests increases, the gap between different algorithms will be smaller and smaller in statistical. As shown in Fig.9, the average latency using ACO-EFPC is a little higher than that using SADR, with the gap is shrinking as the data flows increases, and the latency using ACO-EFPC is lower than that using SADR while 180 the data flow requests arrive. Due to the slow convergence, the latency using Dijkstra algorithm is always higher than other algorithms.

2) THE AMOUNT OF BANDWIDTH FRAGMENTATION

As mentioned above, we mainly focus on the data flows in ISTN, and do not include the data flows transported in the terrestrial networks. The utilization and efficiency of the terrestrial networks may be influenced by the data flows transported only in the terrestrial networks. Therefore, some indicators are much more meaningful to consider the satellite networks instead of the whole ISTN. While the number of data flow requests is 150, the total amount of bandwidth fragmentation in satellite network using different algorithms are illustrated in Fig. 10. Although the total amount of bandwidth fragmentation of the services using ACO-EFPC

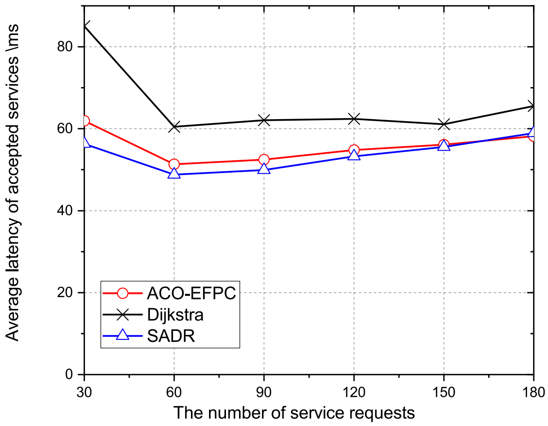

FIGURE 9. The average latency of different number of data flow requests.

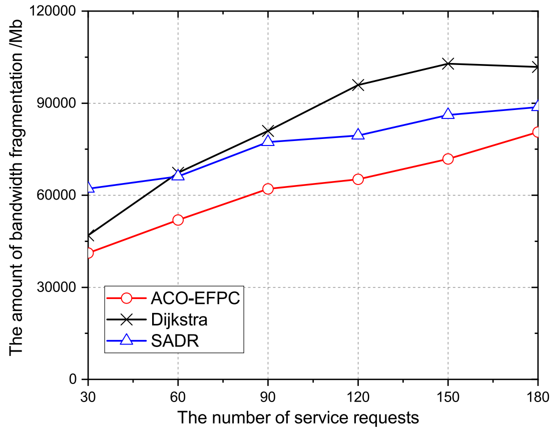

FIGURE 11. The amount of bandwidth fragmentation of different number of data flow requests in satellite networks and STLs.

did not show an advantage at the beginning time intervals, it is lower than that using other algorithms in subsequent time intervals. In general, the average bandwidth fragment using ACO-EFPC is 18.24% less than that using the Dijkstra algorithm whereas the value is 8.22% compared to the SADR algorithm.

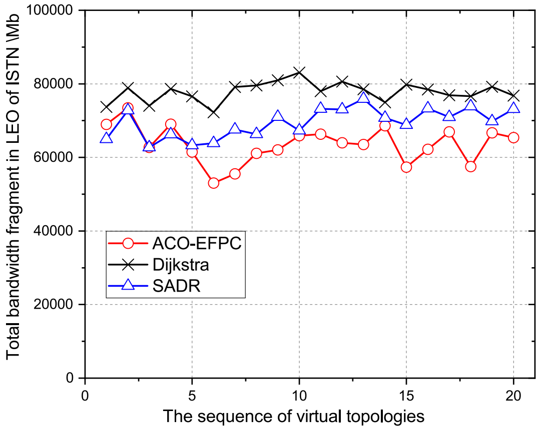

FIGURE 10. The amount of bandwidth fragmentation of 150 data flow requests in satellite networks and STLs.

Fig. 11 illustrates the total bandwidth fragmentation of satellite network and STLs of ISTN under the different numbers of data flow requests. Dijkstra algorithm performs better than SADR when the network load is low, but the amount of bandwidth fragmentation grows fast as the network load increases. The ACO-EFPC and SADR are relatively slow to grow. No matter how much of the network loads, the total amount of bandwidth fragmentation using ACO-EFPC is much lower than that using other algorithms in statistically (23.58% lower than Dijkstra algorithm and 19.78% lower than SADR in average).

3) TOTAL ELASTIC BANDWIDTH

As the elastic data flow model we talked above, the total allocated bandwidth is decided by the elastic bandwidth, a lower elastic bandwidth means a higher degree of aggregation, which need less total bandwidth. The amounts of elastic bandwidth using different algorithms under 150 data flows are shown in Fig. 12. Due to the restriction of MSB overflow factor, ACO-EFPC algorithm has the least elastic bandwidth in any virtual topology, whereas the Dijkstra algorithm performs the worst in elastic bandwidth at any time interval. The total elastic bandwidth using ACO-EFPC is 26.22% less than that using Dijkstra algorithm and 16.32% less than that using SADR on average.

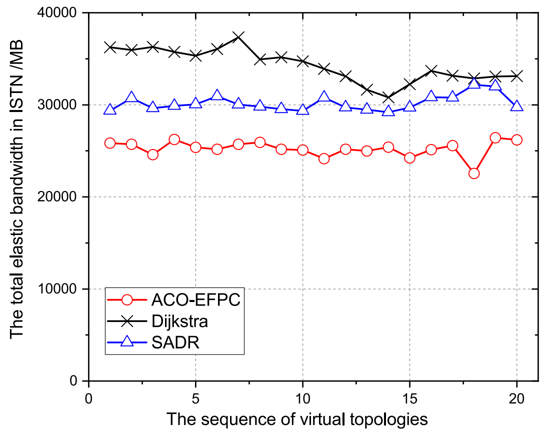

FIGURE 12. The elastic bandwidth of 150 flow requests in satellite networks.

Fig. 13 illustrates the average elastic bandwidth of the data flows in the satellite network of ISTN under the different numbers of flow requests. Similar with wavelength fragmentation, the Dijkstra algorithm perform well at the beginning, but it quickly deteriorated as the network load increased. The performance of ACO-EFPC on elastic bandwidth is better than other algorithms under any network load. The elastic bandwidth in satellite networks of ISTN using ACO-EFPC is 10.23% less than that using Dijkstra, and 9.23% less than that using SADR.

4) LOAD BALANCING

We use the variance of the number of allocated wavelength to evaluate the performance of the algorithms on load balancing.

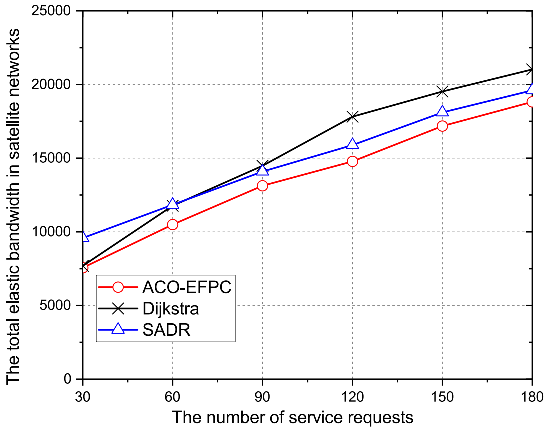

FIGURE 13. The elastic bandwidth of different number of data flow requests in satellite networks.

As the ISLs is based on WDM, it is much more meaningful to use the variance of the number of allocated wavelength than the variance of the allocated bandwidth. Fig. 14 illustrates the variance of the number of the allocated wavelength of the links in the satellite networks of the ISTN under 150 data flows. The results show the Dijkstra algorithm is the worst one, while the SADR and ACO-EFPC perform at the same level, but the ACO-EFPC is better at most times. Statistically, the variance of the traffic using ACO-EFPC is 16.72% lower than that using Dijkstra, and 3.65% lower than that using SADR.

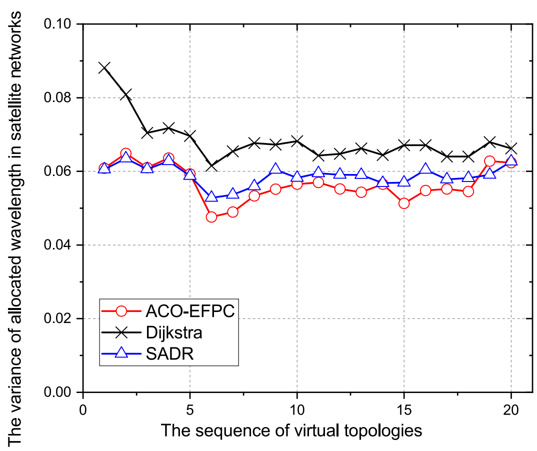

FIGURE 14. The variance of the number of the allocated wavelength of 150 data flow requests in satellite networks.

Meanwhile, Fig. 15 shows the variance of the allocated wavelength in satellite networks of ISTN under the different network loads. When the network is lightly loaded, the SADR and the Dijkstra algorithm have the same performance on load balancing. As network load increases, the variance of the allocated wavelength of Dijkstra increases more than SADR. The ACO-EFPC and SADR almost coincide when the network is under moderate load, and ACO-EFPC performs better than other algorithms when the network load is high. The variance of the allocated wavelength using ACO-EFPC is 16.55% lower than that using Dijkstra and 3,49% lower than that using SADR. The result shows the ACO-EFPC algorithm has a better performance on load balancing.

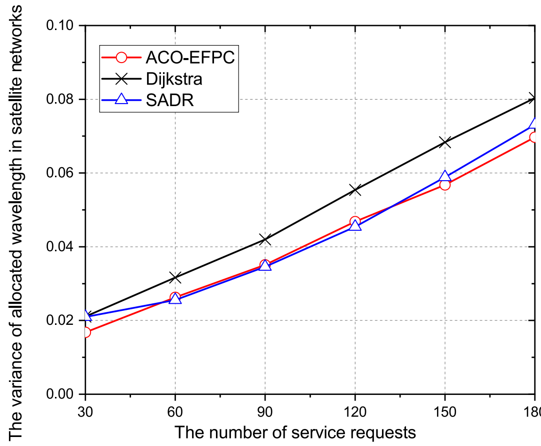

FIGURE 15. The variance of the allocated wavelength of different number of data flow requests in satellite networks.

In general, the ACO-EFPC has got an advantage in bandwidth fragment, bandwidth utilization, and load balancing, only with little inferior in latency while network load is high. For any fixed number of data flows, the indicators are changed over time. Though we do not simulate the ACO-EFPC in every time slice, the corresponding indicators are statistically advantageous.

VI. CONCLUSION

In this paper, faced with the two challenges as independent routing and potential dynamic links in ISTN, We first describe the typical ISTN scenario and the SDN-based unified management architecture, and proposed the elastic data flow model and unified routing model with the considerations as latency, wavelength fragment, allocated bandwidth, and load balancing in the integrated network. Then ACO-EFPC was presented to solve the unified routing problem. As the novelty of the data flow model and the unified routing problem model, ACO-EFPC contains two sub-algorithm for new coming data flows and forwarding data flows respectively. The numerical simulation shows that the capacity of ISTN under the end-to-end routing mechanism has been significantly improved, and the proposed algorithm owns the advantages on wavelength fragment, allocated bandwidth, and load balancing with little additional latency while network load is high.

REFERENCES

[1] Z. Tang, B. Zhao, W. Yu, Z. Feng, and C. Wu, "Software defined satellite networks: Benefits and challenges," in *Proc. IEEE Comput., Commun. IT Appl. Conf.*, Oct. 2014, pp. 127–132.

[2] T. Li, H. Zhou, H. Luo, and S. Yu, "SERvICE: A software defined framework for integrated space-terrestrial satellite communication," *IEEE Trans. Mobile Comput.*, vol. 17, no. 3, pp. 703–716, Mar. 2018.

[3] R. Ferrús *et al.*, "SDN/NFV-enabled satellite communications networks: Opportunities, scenarios and challenges," *Phys. Commun.*, vol. 18, pp. 95–112, Mar. 2016.

[4] B. Feng *et al.*, "HetNet: A flexible architecture for heterogeneous satellite-terrestrial networks," *IEEE Netw.*, vol. 31, no. 6, pp. 86–92, Nov./Dec. 2017.
[5] NetWorld2020s—SatCom WG, "The role of satellites in 5G," White Paper, Jul. 2014.
[6] Z. Qu, G. Zhang, H. Cao, and J. Xie, "LEO satellite constellation for Internet of Things," *IEEE Access*, vol. 5, pp. 18391–18401, 2017.
[7] N. Zhang, S. Zhang, P. Yang, O. Alhussein, W. Zhuang, and X. S. Shen, "Software defined space-air-ground integrated vehicular networks: Challenges and solutions," *IEEE Commun. Mag.*, vol. 55, no. 7, pp. 101–109, Jul. 2017.
[8] F. Alagoz, O. Korcak, and A. Jamalipour, "Exploring the routing strategies in next-generation satellite networks," *IEEE Wireless Commun.*, vol. 14, no. 3, pp. 79–88, Jun. 2007.
[9] J. Foust, "The return of the satellite constellations," *Space Rev.*, Mar. 2015. [Online]. Available: http://www.thespacereview.com/article/2716/1
[10] LeoSat. (2018). *Corporate Overview*. [Online]. Available: http://leosat.com/media/1108/leosat-corporate-overview.pdf
[11] Iridium Commun. Inc. (2018). *Iridium Global Network*. [Online]. Available: https://www.iridium.com/network/globalnetwork
[12] R. Ferrus, O. Sallent, T. Ahmed, and R. Fedrizzi, "Towards SDN/NFV-enabled satellite ground segment systems: End-to-End traffic engineering use case," in *Proc. IEEE Int. Conf. Commun. Workshops*, May 2017, pp. 888–893.
[13] Y. Wang, Y. Xu, Y. Zhang, and P. Zhang, "Hybrid satellite-aerial-terrestrial networks in emergency scenarios: A survey," *China Commun.*, vol. 14, no. 7, pp. 1–13, Jul. 2017.
[14] Y. Miao, Z. Cheng, and Z. Cui, "Software defined integrated satellite-terrestrial network: A survey," in *Proc. Int. Conf. Space Inf. Netw.* Singapore: Springer, 2016, pp. 16–25.
[15] C. Niephaus, M. Kretschmer, and G. Ghinea, "QoS provisioning in converged satellite and terrestrial networks: A survey of the state-of-the-art," *IEEE Commun. Tuts.*, vol. 18, no. 4, pp. 2415–2441, 4th Quart., 2016.
[16] X. Artiga, J. Nunez-Martinez, A. Perez-Neira, G. J. L. Vela, J. M. F. Garcia, and G. Ziaragkas, "Terrestrial-satellite integration in dynamic 5G backhaul networks," in *Proc. 8th IEEE Adv. Satell. Multimedia Syst. Conf. (ASMS)*, Sep. 2016, pp. 1–6.
[17] G. Wen, Q. Zhang, H. Wang, Q. Tian, and Y. Tao, "An ant colony algorithm based on cross-layer design for routing and wavelength assignment in optical satellite networks," *China Commun.*, vol. 14, no. 8, pp. 63–75, Aug. 2017.
[18] W. Houtian, Z. Qi, X. Xiangjun, T. Ying, and L. Naijin, "Cross-layer design and ant-colony optimization based routing algorithm for low earth orbit satellite networks," *China Commun.*, vol. 10, no. 10, pp. 37–46, Oct. 2013.
[19] Y. Dong, S. Zhao, H. Dan Ran, Y. Li, and Z. Zhu, "Routing and wavelength assignment in a satellite optical network based on ant colony optimization with the small window strategy," *J. Opt. Commun. Netw.*, vol. 7, no. 10, pp. 995–1000, Oct. 2015.
[20] X. Liu, X. Yan, Z. Jiang, C. Li, and Y. Yang, "A low-complexity routing algorithm based on load balancing for LEO satellite networks," in *Proc. IEEE 82nd Veh. Technol. Conf.*, Sep. 2015, pp. 1–5.
[21] H. Cheng, M. Liu, S. Wei, and B. Zhou, "A distributed algorithm for self-adaptive routing in LEO satellite network," in *Proc. Int. Conf. Space Inf. Netw.* Singapore: Springer, 2016, pp. 274–286.
[22] D. Yan, J. Guo, L. Wang, and P. Zhan, "SADR: Network status adaptive QoS dynamic routing for satellite networks," in *Proc. IEEE 13th Int. Conf. Signal Process. (ICSP)*, Nov. 2016, pp. 1186–1190.
[23] Z. Xing *et al.*, "Dynamic and efficient networking of satellite optical communication for multiple provisional traffic," in *Proc. Asia Commun. Photon. Conf.*, 2017, Paper Su2A64.
[24] K. Maine, C. Devieux, and P. Swan, "Overview of IRIDIUM satellite network," in *Proc. WESCON*, Nov. 1995, p. 483.
[25] F. J. Dietrich, P. Metzen, and P. Monte, "The Globalstar cellular satellite system," *IEEE Trans. Antennas Propag.*, vol. 46, no. 6, pp. 935–942, Jun. 1998.
[26] T. Dong and G. Shen, "Traffic grooming for IP over WDM optical satellite networks," in *Proc. Int. Conf. Opt. Commun. Netw. (ICOCN)*, Nov. 2014, pp. 1–6.
[27] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numer. Math.*, vol. 1, no. 1, pp. 269–271, 1959.
[28] M. R. Garey and D. S. Johnson, "*Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco, CA, USA: Freeman, 1979, pp. 90–91.
[29] (2018). *Open vSwitch Documentation Contents*. [Online]. Available: http://docs.openvswitch.org/en/latest/contents/
[30] MPLS-TP. (2017). *OpenFlow Protocol Extensions for SPTN, Version 1.0, The Open Networking Foundation*. [Online]. Available: https://www.opennetworking.org/wp-content/uploads/2017/07/MPLS-TP-OpenFlow-Protocol-Extensions-for-SPTN-1-0.pdf
[31] M. Karakus and A. Durresi, "Quality of service (QoS) in software defined networking (SDN): A survey," *J. Netw. Comput. Appl.*, vol. 80, pp. 200–218, Feb. 2017.

QIZE GUO received the B.E. degree in communication engineering from Information Engineering University, China, in 2009. He is currently pursuing the Ph.D. degree with the Beijing University of Posts and Telecommunications, China.

His research interests are mainly in the field of integrated satellite-terrestrial network, optical transport network, franthaul/backhaul network, and software-defined network.

RENTAO GU received the Ph.D. degree from the Beijing University of Posts and Telecommunications (BUPT), China.

He is currently an Associate Professor with BUPT. His current research interests include optical network and intelligent information processing. He is a Senior Member of the China Institute of Communications.

TAO DONG received the B.S. and Ph.D. degrees in electronics science and technology from the Beijing Institute of Technology, in 1999 and 2004, respectively.

He is currently a Professor with the State Key Laboratory of Space-Ground Integrated Information Technology, Beijing Institute of Satellite Information Engineering, China. His research interests include optical phased array, free space optical communication, and space optical networks.

JIE YIN received the Ph.D. degree from the Beijing University of Posts and Telecommunications, China.

He is currently a Senior Engineer with the State Key Laboratory of Space-Ground Integrated Information Technology, Beijing Institute of Satellite Information Engineering. His research interests are in the fields of satellite-terrestrial network, free space optical communication, and microwave photonics.

ZHIHUI LIU received the M.S. degree in computational mathematics from Xiangtan University, Xiangtan, in 2012, and the Ph.D. degree in signal and information processing from the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China.

She is currently an Engineer with the State Key Laboratory of Space-Ground Integrated Information Technology, Beijing Institute of Satellite Information Engineering. Her research interests include cognitive radio networks, green communication, wireless communication, and satellite communication networks.

LIN BAI is currently a Professor with the Beijing University of Posts and Telecommunications, China. Her research interests are primarily in the areas of information management, information theory, and network technology.

YUEFENG JI received the Ph.D. degree from the Beijing University of Posts and Telecommunications (BUPT), China.

He is currently a Professor with BUPT. His research interests are primarily in the area of broadband communication networks and optical communications, with emphasis on key theory, and realization of technology and applications. He is a fellow of the China Institute of Communications.