Intelligent Automation & Soft Computing

Tech Science Press

DOI: 10.32604/iasc.2023.038477
Article

Satellite-Air-Terrestrial Cloud Edge Collaborative Networks: Architecture, Multi-Node Task Processing and Computation

Sai Liu[1], Zhenjiang Zhang[1,*], Guangjie Han[2] and Bo Shen[1]

[1]School of Electronic Information Engineering, Beijing Jiaotong University, Beijing, 100044, China
[2]School of Internet of Things Engineering, Hohai University, Nanjing, 211100, China
*Corresponding Author: Zhenjiang Zhang. Email: zhangzhenjiang@bjtu.edu.cn
Received: 14 December 2022; Accepted: 12 April 2023; Published: 11 September 2023

Abstract: Integrated satellite-terrestrial network (ISTN) has been considered a novel network architecture to achieve global three-dimensional coverage and ultra-wide area broadband access anytime and anywhere. Being a promising paradigm, cloud computing and mobile edge computing (MEC) have been identified as key technology enablers for ISTN to further improve quality of service and business continuity. However, most of the existing ISTN studies based on cloud computing and MEC regard satellite networks as relay networks, ignoring the feasibility of directly deploying cloud computing nodes and edge computing nodes on satellites. In addition, most computing tasks are transferred to cloud servers or offloaded to nearby edge servers, the layered design of integrated satellite-air-terrestrial architecture and the cloud-edge-device cooperative processing problems have not been fully considered. Therefore, different from previous works, this paper proposed a novel satellite-air-terrestrial layered architecture for cloud-edge-device collaboration, named SATCECN. Then this paper analyzes the appropriate deployment locations of cloud servers and edge servers in ISTN, and describes the processing flow of typical satellite computing tasks. For computing resource allocation problems, this paper proposed a device-edge-cloud Multi-node Cross-layer Collaboration Computing (MCCC) method to find the optimal task allocation strategy that minimizes the task completion delay and the weighted system energy consumption. Furthermore, the approximate optimal solutions of the optimization model are obtained by using successive convex approximation algorithm, and the outstanding advantages of the proposed method in reducing system energy consumption and task execution delay are verified through experiments. Finally, some potential issues and directions for future research are highlighted.

Keywords: Device-edge-cloud collaboration; ISTN; MEC; task computation; resource allocation

1 Introduction

With the development of the fifth-generation (5G) communication network, the world is witnessing the arrival of the Internet of Things (IoT) era. Sensors, vehicles, tablets, wearable devices and more have joined the network, spawning a host of technologies and applications. For example, smart city, intelligent transportation [1], coastal monitoring [2], intelligent agriculture [3], augmented reality (AR)/virtual reality (VR), intelligent healthcare [4] and so on are developing rapidly. Global anytime and anywhere network access, large-scale machine-type communication and ultra-reliable low latency communication have become the development trend of the 6G network in the future [5]. Especially as the IoT continues to evolve, the number of IoT devices will exceed 24 billion by 2030 [6], where ubiquitous communication coverage is critical to support widely distributed devices. However, due to economic and technical considerations, traditional cellular base stations (BSs) are difficult to deploy in remote areas with complex terrain and sparse population [7]. Digital services and computing based on terrestrial networks are limited to densely populated areas, while uninhabited areas such as deep space, ocean and desert are still blank areas. In addition, the current terrestrial network infrastructure is vulnerable to natural disasters such as earthquakes and hurricanes. In recent years, satellites with high throughput and low latency have attracted wide attention. Satellite communication is no longer just a supplement to terrestrial networks. Combining the advantages of satellite networks and terrestrial networks to construct an ISTN architecture is an important trend of future network development. Based on the terrestrial cellular mobile network, the ISTN architecture combining the characteristics of wide coverage, flexible deployment and efficient broadcasting of broadband satellite communication, could realize the global three-dimensional coverage through the deep integration of multiple heterogeneous networks. As computing follows the network, collaborative satellite-air-terrestrial computing will integrate satellite systems, aerial networks, terrestrial communication, cloud computing and MEC, becoming an emerging computing architecture that expands the space for digital services. In this sense, there will be new opportunities to support ubiquitous IoT applications.

Satellite communication provides a direct solution to the coverage problem with its extensive coverage capability. With the increase of communication demand and the progress of communication technology, utilizing satellite constellation to achieve global coverage has become a research hotspot in academia and industry. A comprehensive summary of the existing ISTN architecture is presented in [8]. The ISTN architecture can extend network coverage based on satellite backhaul transmission and improve network reliability by introducing terrestrial relay into the satellite network. In [9], the author put forward three basic satellite-terrestrial cooperation models: Model X, L and V, and points out that ISTN is the combination of these three basic cooperation models. Reference [10] described a potential architecture that combines extended space networks, the Internet and mobile wireless networks to provide comprehensive services and global network access. The research work of SATIN in recent years is reviewed from the aspects of network design, resource allocation, performance analysis and optimization. Several existing network architectures are discussed, some technical challenges and future development directions are pointed out in [11]. In [12], the authors studied the fusion of satellite network and the 5G network, pointing out that the adoption of software defined network (SDN) and network function virtualization (NFV) technology in satellite communication domain is the key element to realize this fusion. The authors of [13,14] applied SDN technology to ISTN, which played a significant role in promoting network design. In [15], the authors studied the cooperative transmission problem in ISTN.

Mobile devices with limited battery power and computing ability cannot satisfy the computing requirements of emerging IoT applications. In most cases, mobile users can offload computing tasks to cloud data centers with abundant computing resources to compensate for the deficiency of computing and storage resources on devices. However, cloud data centers are always far from mobile users, resulting in a large delay in the processing of mobile services, which is difficult to achieve the millisecond end-to-end delay standard for uRLLC applications. As an emerging network technology, edge computing [16,17] has the potential to improve quality of experience and reduce redundant network traffic. By extending cloud computing platforms to the edge, the network can provide users with multiple layers of heterogeneous computing resources that can be accessed from anywhere in the world. Currently, the combination of MEC and ISTN is also attracting considerable attention [18–20]. With the assistance of satellites and aerial vehicles, the authors in [18] proposed a SATIN edge/cloud computing architecture for IoT systems. The authors of [19] proposed satellite mobile edge computing (SMEC), which integrates the MEC technique in ISTN-based mobile communication systems. Moreover, a cooperative computation offloading method for the SMEC scenario is presented. In [20], the authors developed a software-defined ISTN to manage and coordinate the network, caching, and computing resources, and design a deep learning approach to solve joint resource allocation problems.

Although it has been realized that MEC plays a key role in improving the service of the 6G integrated satellite-terrestrial network, most of the proposed works regard satellite networks as relay networks, that is, the user's computing tasks are transmitted to the terrestrial cloud data centers or edge servers for processing through the satellite relay. The feasibility of deploying cloud and edge computing resources directly on satellites is largely ignored. By directly deploying computing resources of different sizes on satellites, the QoS of users with massive computing requirements such as audio and video can be further improved through the collaborative processing of miniature cloud computing nodes and edge computing nodes. In addition, the existing research work only considers the basic model of MEC enabling ISTN, lacking the design of cloud-edge-device integration architecture. The multiple layers and synergies of computing resources have also not been fully considered.

Therefore, different from previous works, this paper makes a comprehensive study on satellite-air-terrestrial cloud edge collaborative network which considers the layered utilization of cloud computing and edge computing resources to improve resource utilization and service guarantee ability of ISTN. The main contributions of this paper are:

- A deep fusion layered satellite-air-terrestrial cloud edge collaborative network (SATCECN) architecture is proposed, and its main network structure and application functions are discussed.
- The roles of cloud computing and MEC in SATCECN are analyzed, and the possible deployment locations of cloud servers and edge servers in ISTN are given. On this basis, the processing flow of various typical tasks are described in detail.
- Aiming at low delay and high energy efficiency of task processing, optimization problem expressions are constructed. A device-edge-cloud Multi-node Cross-layer Collaborative Computing (MCCC) method for SATCECN is proposed, and its outstanding advantages in reducing system energy consumption and task execution delay are verified by experiments.

The remainder of the paper is organized as follows. Section 2 introduces the integrated satellite-air-terrestrial communication network architecture and its functional components in detail. Section 3 discusses the deployment principles of cloud servers, edge servers, and the processing flow of typical tasks in SATCECN. Following that, this paper introduces the multi-node cross-layer collaborative computing method in Section 4. The advantages of MCCC in reducing system energy consumption and task execution delay are verified by conducting experiments in Section 5. Finally, the paper is summarized and some potential problems and directions for future research are pointed out.

2 Integrated Satellite-Air-Terrestrial Communication Network

This section first analyzes the characteristics of air networks and satellite networks. Then, combing the advantages of them with mature terrestrial networks, an integrated satellite-air-terrestrial communication network architecture is proposed. And the functions of its key parts are explained.

GEO satellites operate in a fixed orbit at an altitude of 36000 kilometers. Due to the high orbital altitude, it can provide the most comprehensive coverage, but the propagation delay and signal attenuation are correspondingly large. LEO satellites typically orbit at an altitude from 500 to 2000 kilometers. At lower altitude, the propagation delay and signal attenuation are much less than those of GEO and MEO satellites. In addition, the altitude of LEO satellites is not fixed and the orbit resources are sufficient. Therefore, it is feasible to achieve fast global broadband access using LEO satellite constellations, which is very promising for supporting applications with ultra-low latency and ultra-high reliability requirements in the 6G network. MEO satellites generally operate at an orbit altitude from 2000 to 36000 km, with smaller propagation delay and signal attenuation than GEO satellites, and larger coverage than LEO satellites, which have the advantages of both GEO and LEO satellites. Since MEO satellites have relatively limited energy supply and computing capacity, they can act as relay nodes to provide relay services for data transmission.

Air network typically consists of balloons, aircraft and unmanned aerial vehicles (UAVs) deployed at altitudes below 30 km. Using solar energy, balloons can provide continuous service to specific areas for several years [21]. At the same time, due to mobility and flexibility, high-altitude platforms like balloons are much cheaper to deploy and can provide tens of kilometers of rural coverage in a quasi-stationary manner. UAVs have smaller size and higher mobility which are easy to deploy [22]. However, due to battery limitations, the operation time is relatively short. Low-altitude platforms such as UAVs can enhance transmission in hotspots congested by heavy traffic [23,24]. In addition, with the growing development of IoT, UAVs can serve as aerial BSs or relay stations to provide temporary wireless access to IoT devices in rural and remote areas.

As shown in Fig. 1, the integrated satellite-air-terrestrial communication network is mainly composed of space layer, air layer and terrestrial layer. The three layers can cooperate with each other in different application scenarios. By integrating the three-layer heterogeneous network, a full-domain three-dimensional coverage broadband wireless network with deep fusion can be constructed.

Space Network: The GEO, MEO satellites, LEO satellite constellations and their corresponding terrestrial stations, antenna array fields, network control centers and other infrastructure constitute the space network. GEO satellites have a large coverage area and high-speed connection between space and earth, which can be used for data broadcasting service of popular network requests. One or more GEO satellites with powerful onboard computing and storage capabilities can be used as spatial data centers, with functions such as data storage, processing, routing and network management. The spatial data center can process data in real time, thereby reducing the delay and link load caused by the space-earth communication. In addition to being relay satellites, MEO satellites can also provide voice and low-rate data services to users worldwide. By connecting with terrestrial gateway stations with antenna array fields and terrestrial terminal stations with limited capacity, LEO satellites can respectively provide mobile communication services for dense users distributed in the same area and sparse users located in remote villages, mountains and oceans.

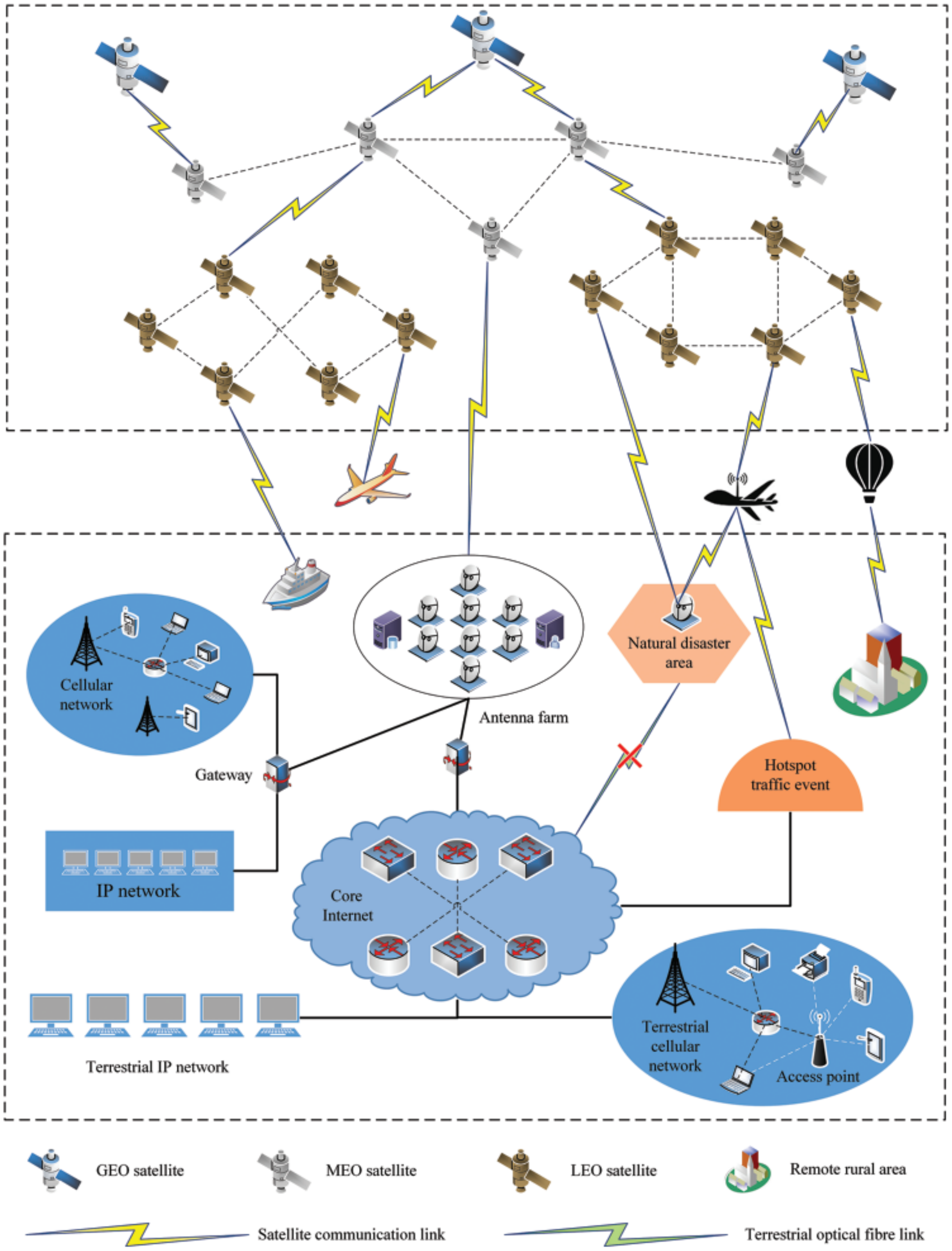


Figure 1: The integrated satellite-air-terrestrial communication network

Air Network: Airborne mobile system for information collection, transmission and processing based on aircraft. UAVs, airships and balloons serve as aerial platforms to provide broadband wireless communication to complement terrestrial networks. Compared with the base stations in terrestrial network, the air network has the characteristics of low cost, convenient to deploy and large coverage, and can provide regional wireless access services.

Terrestrial Network: It is mainly composed of terrestrial communication systems, such as cellular networks, MANET [25], Global Interoperability Microwave Access, Wireless Local Area Networks (WLANs), which can provide users with high-rate data transmission.

Despite the great potential of the integrated satellite-air-terrestrial collaborative network, the increasing QoS requirements of users, such as high data rate, low communication delay and processing energy consumption may limit its widespread application. Especially in recent years, with the rapid development of intelligent terminals, many new computing and energy intensive applications have emerged, such as Automatic Speech Recognition (ASR), Games, Multimedia coding/decoding and Self-driving. Although cloud computing can be used to complete these tasks, it will be affected by data transmission delay or unreliable delay jitter in wide area network, so that the QoS requirements of users cannot be met. In addition, there are challenges such as scalability, security and reliability in the integrated satellite-air-terrestrial collaborative network. The collaborative computing of edge, cloud and intelligent terminals can effectively address the above challenges, which has become an important development trend in the future.

3 Cloud-Edge Collaborative Task Processing

This section analyzes and discusses the principles of deploying cloud computing resources and edge computing resources in the integrated satellite-air-terrestrial communication network respectively from the perspectives of user requirements and service capability. Then the SATCECN architecture is proposed. Furthermore, according to the computing ability and energy situation of different platforms, the typical task cooperative processing flow is introduced from the perspectives of user density and task type.

3.1 Computing Resource Deployment

Due to the outstanding computing and storage advantages of cloud computing, the current communication networks are widely based on cloud processing to provide services [26]. However, digital services based on terrestrial networks are limited to densely populated areas. Unpopulated areas such as deep space, ocean and desert are still blank areas of service. In the SATCECN, space-air-terrestrial collaborative computing will integrate satellite systems, high and low altitude flight platforms, and terrestrial communications, becoming an emerging computing architecture to expand the space of digital services. The satellite network and terrestrial Internet will be seamlessly combined to form a universal Internet of space and earth, and satellite and terrestrial systems will become new computing nodes relying on cloud computing capabilities. Despite the continuous improvement of satellite onboard capacity, the data storage and data processing capacity of a single satellite is still limited due to the limitations of on-board physical space and energy. Therefore, the deployment of cloud computing in the space network is limited compared to terrestrial networks. The deployment locations of cloud computing in the integrated satellite-air-terrestrial communication network can generally be classified into the following two types:

1) Deployed on the terrestrial data center

Data centers located in the terrestrial backbone network have abundant computing and storage resources and can process computing tasks generated by various user devices in the network.

2) Deployed on the satellites with strong payload capacity

Since the geosynchronous orbit satellites can fly around the earth without additional power after gaining the first cosmic speed, they have more energy to be used for computation and storage compared with MEO satellites and LEO satellites. GEO satellites can be used as cloud platforms in the space network to supplement additional computing resources for satellite edge nodes. However, the orbital altitude of GEO satellites inevitably brings large propagation delay and signal attenuation. For delay-sensitive tasks, computing resources need to be closer to the data source. In order to balance the delay requirement and computing resource requirement, multiple satellite computing nodes located in LEO satellite constellation can form a distributed cloud computing capability to provide on-demand services.

Compared to cloud computing, MEC is an open platform that can be flexibly deployed at different locations in ISTN to meet different service requirements. The deployment locations of MEC in the integrated satellite-air-terrestrial communication network can be generally classified into the following three types.

1) Deployed at base stations in the terrestrial network

Under normal circumstances, MEC servers can be deployed at terrestrial base stations near the user. Computing tasks can be directly offloaded to the MEC server at the base station near the user, and processed nearby. This deployment mode requires available base stations nearby and is more suitable for dense user scenarios. For sparse users connected to satellites via terrestrial terminal station, there is no additional computing resources. It is also not economically cost-effective to deploy dedicated MEC servers for small and dispersed groups of users.

2) Deployed on satellites in low earth orbit

MEC servers can also be deployed on LEO satellites. In this case, all terrestrial communication devices within satellite coverage can benefit from the offloading service. This deployment mode reduces traffic between the satellite and the terrestrial backbone network. The delay of the proposed scheme is relatively high compared to the deployment on the terrestrial base station side, but it still has a significant improvement in the delay compared to sending requests through the terrestrial station to the remote cloud.

3) Deployed at the terrestrial gateway station

The MEC servers are deployed in the terrestrial gateway station to provide more computing resources to all users connected to the gateway within satellite coverage. Thus, a large amount of data transmission between the satellite users and the remote cloud is avoided. Compared with other above offloading methods, this deployment method leads to relatively high service delay due to the extra relay of LEO satellite links, but it is more practical in implementation and maintenance.

Fig. 2 shows the computing resources deployment location and the typical cooperative structure of cloud, edge and terminals in SATCECN. Terrestrial facilities such as MEC platforms, data centers, and gateways communicate with user devices through cellular networks and receive satellite computing tasks through antenna arrays at the terrestrial gateway station. Data centers are rich in computing and storage resources and can process computing tasks generated by various user devices in the network. The gateways implement routing, forwarding, traffic control, unified network management and orchestration.

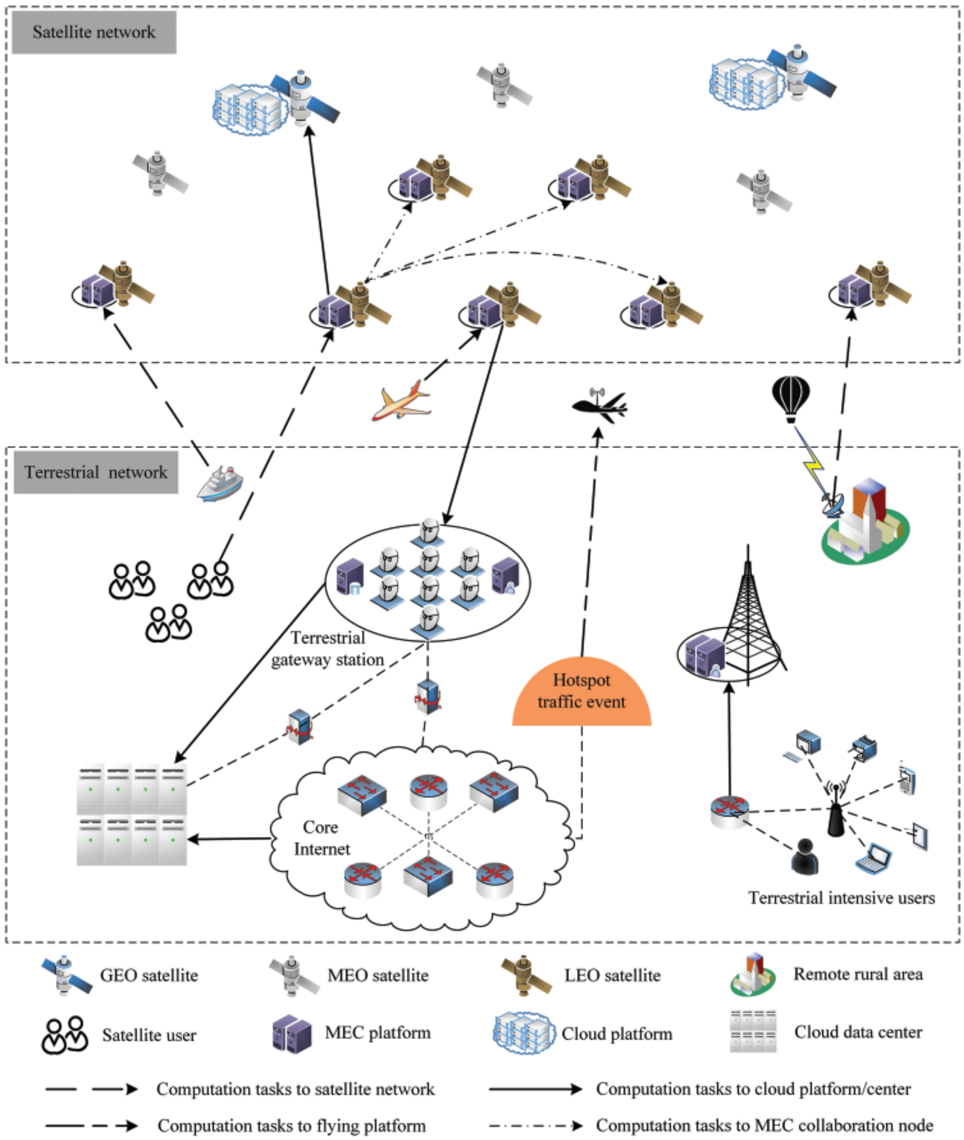


Figure 2: Satellite-air-terrestrial cloud edge collaborative network and resources deployment

3.2 Task Processing

According to user density, the application scenarios of the SATCECN can be divided into intensive user scenarios and sparse user scenarios. In intensive user scenarios, the 4G/5G services can be provided by deploying cellular towers and eNodeB to build a terrestrial cellular network. Computing tasks can be offloaded to the MEC platforms at the base station side for processing, instead of being transmitted to the data centers through the core Internet. For sparse users distributed in high-altitude platforms, aircraft, isolated islands, rural areas and other areas, the terrestrial cellular mobile networks cannot cover, and the communication, computing, cache and storage capacities of user terminals are very limited. Therefore, these user terminals can communicate through satellites and offload their computing tasks directly to the MEC platforms on LEO satellites. They can also offload their computing tasks to the cloud platforms on GEO satellites or the terrestrial data centers for processing through MEO satellites relay.

According to different requirements, tasks can be divided into ordinary tasks, delay sensitive tasks and large-scale communication tasks. For ordinary computing tasks generated by users, the user first checks whether local resources can meet computing requirements. If not, the tasks will be offloaded to the edge computing nodes in the satellite network or the terrestrial network depending on the user's environment. When a corresponding edge computing node receives the computing tasks, if it is in a busy state, it will send the computing tasks to the cloud computing platforms in the satellite network or the terrestrial data centers for processing. Otherwise, it will decide whether to cooperate with the surrounding edge computing nodes to process the computing tasks, which depends on the collaborative task scheduling strategy of the edge computing network. For delay-sensitive computing tasks, the computing results are expected to be returned to the target users in the shortest time, so they are generally not sent to distant data centers for processing. After delay-sensitive tasks are generated, the user first evaluates the computing resources required by the tasks, and then obtains the idle computing resources information of MEC platforms nearby and LEO satellites during the illumination time from the gateway control center. If the demand is met, the tasks will be distributed to them. Otherwise, the nearest LEO satellite will distribute the tasks to other LEO satellites or spatial data centers with sufficient idle computing capacity through high-speed inter-satellite laser link. After the calculation is completed, the results will be sent to the LEO satellite within the illumination time of the target user, and then forwarded to the target user. Large-scale communication tasks, mainly aimed at multicast and broadcast tasks of users in a wide geographical distribution range. Satellite networks have extensive coverage capabilities and are naturally suitable for broadcast and multicast services. Common contents required by most users are transmitted via GEO/MEO satellites broadcasting. The public contents needed by small-scale and dispersed users are transmitted by GEO/MEO satellites multicast according to their coverage area. The unique content required by each user is transmitted unicast by the terrestrial broadcast servers. Satellite and terrestrial networks cooperate to publish contents, improving QoS of users and resources sharing efficiency.

In general, the entire network architecture greatly reduces the service response time by deploying cloud and edge computing resources closer to the users in ISTN. The collaboration between cloud computing nodes, edge computing nodes and user terminals further accelerates the task processing speed and provides users with better service experience.

4 Cooperative Computing

With the increasing demand for low latency and high energy efficiency in task processing, it is of great significance to distribute computing tasks to multiple available service nodes on space, ground and flight platforms for joint computing to improve the network performance of ISTN. However, the optimal allocation of multi-node and cross-layer computing resources in the SATCECN faces many technical challenges. For example, computing nodes in the satellite network (such as LEO satellites) usually move fast, and their high mobility makes the traditional resources allocation algorithm need to update rapidly and dynamically according to the network status, which puts forward higher requirements for network information interaction.

This section provides a Multi-node Cross-layer Collaborative Computing (MCCC) method. It can solve the computing resources allocation problem in the SATCECN, overcome the storage resources occupation problem of high-dimensional data, and has a good generalization ability to improve computing efficiency. The implementation steps of this method are as follows:

(1) The satellite user terminal devices (local compute nodes) obtain the computing resources information of each edge collaboration node and cloud collaboration node, the wireless transmission information between each computing node through the nearest satellite edge node within the visible range, and then determine the edge collaboration computing nodes and cloud collaboration nodes from the nodes with computing capacity.
(2) According to the task execution status information, the resources information of each compute node and the wireless transmission information between compute nodes, the expressions of the optimization problems are constructed to minimize the system weighted energy consumption and task execution delay. Then the system energy-delay model is established.
(3) Determine the resources allocation strategy of each compute node according to the solved energy-delay model.

The execution status information of the task includes the data volume to be executed, the execution delay requirement, the execution status of the task and the visibility of the terrestrial/satellite compute nodes to the task at any time within the decision time range of the local compute node. The wireless transmission information between the service nodes includes satellite-terrestrial data transmission rate δ_{tr}^{dc}, δ_{tr}^{ds} and intersatellite data transmission rate δ_{tr}^{cs}, δ_{tr}^{ss}. Now, let D_v be the total data volume of the computing task v, R_v be the size of the computed results of task v. Define β is the ratio of the data volume executed by the user terminal to D_v, γ_i is the ratio of the data volume executed by the i th spatial edge collaboration node to D_v. So $(1-\beta-\gamma)$ is the ratio of the data volume executed by the spatial cloud collaboration node to D_v, where $\gamma=\sum_i \gamma_i$. Then the distribution time of task v is given by

$$T_{dis}=\frac{(1-\beta-\gamma)\,D_v}{\delta_{tr}^{dc}}+\frac{\gamma D_v}{\delta_{tr}^{ds}} \tag{1}$$

and the backhaul transmission time of the computed results of task v is given by

$$T_{back}=\frac{(1-\beta-\gamma)\,R_v}{\delta_{tr}^{cs}}+\frac{\gamma R_v}{\delta_{tr}^{ss}}+\frac{R_v}{\delta_{tr}^{ds}} \tag{2}$$

During the execution of task v, the computation delay should be the maximum time of the terminal devices, the spatial cloud collaborative nodes and the spatial edge collaborative nodes to complete the corresponding tasks that are distributed. Where, the time of all the spatial edge collaborative nodes to complete their assigned computing tasks can be written as

$$T_{com}^{i}=\max_{i}\frac{\gamma_i D_v}{\eta_s^i} \tag{3}$$

Hence, the time spent to compute the task v corresponds to

$$T_{com}=max\left\{\frac{\beta D_v}{\eta_d},\ \frac{(1-\beta-\gamma)\,D_v}{\eta_c},T_{com}^{i}\right\} \tag{4}$$

where η_d, η_c, η_s^i represent the volume of data that can be calculated per second by the terminal device, the spatial cloud collaboration node, and the i th spatial edge collaboration node, respectively. Thus, we can write the overall task processing latency as follows:

$$T_{total}=T_{dis}+T_{com}+T_{back} \tag{5}$$

The optimization problem expression for minimizing task execution delay can be written as

$$\min\ T_{total} = (1-\beta-\gamma)\left(\frac{D_v}{\delta_{tr}^{dc}} + \frac{R_v}{\delta_{tr}^{cs}}\right) + \gamma\left(\frac{D_v}{\delta_{tr}^{ds}} + \frac{R_v}{\delta_{tr}^{ss}}\right) + max\left\{\frac{\beta D_v}{\eta_d}, \frac{(1-\beta-\gamma) D_v}{\eta_c}, T_{com}^{i}\right\} + \frac{R_v}{\delta_{tr}^{ds}} \quad (6)$$

$s.t.\ 0 \le \beta \le 1;$

$0 \le \gamma_i \le 1;$

$\beta + \gamma \le 1$

Define e_v^d, e_v^c and $e_v^s(i)$ as the energy consumption caused by the independent execution of task v by the local compute node, the spatial cloud collaboration node and the i th spatial edge collaboration node, respectively. e_{tr}^{dc} and $e_{tr}^{ds}(i)$ refer to the transmission energy consumption of the terminal device caused by the user sending all the task data to the spatial cloud collaboration node and the i th satellite edge collaboration node, respectively. e_{tr}^{cs} and $e_{tr}^{ss}(i)$ represent the energy consumption of the spatial cloud collaboration node and the i th spatial edge collaboration node transmitting the calculated result R_v to the backhaul satellite, respectively. e_{tr}^{sd} is the energy consumption generated by the backhaul satellite to transmit the calculation results back to the target user. Then the optimization problem expression for minimizing the system energy consumption can be written as

$$\min\ E = \alpha\left[\beta e_v^d + (1-\beta-\gamma) e_{tr}^{dc} + \sum \gamma_i e_{tr}^{ds}(i)\right] \quad (7)$$

$$+ (1-\alpha)\left[(1-\beta-\gamma)\left(e_v^c + e_{tr}^{cs}\right) + \sum \gamma_i\left[e_v^s(i) + e_{tr}^{ss}(i)\right] + (1-\beta) e_{tr}^{sd}\right]$$

$s.t.\ 0 \le \alpha \le 1;$

$0 \le \beta \le 1;$

$0 \le \gamma_i \le 1;$

$\beta + \gamma \le 1$

where α represents the weight of the terminal device's energy consumption in system energy consumption cost E, and $(1-\alpha)$ represents the weight of the total energy consumption of all the spatial collaboration nodes in the system energy consumption cost E.

To solve the above non-convex problem (6) and (7), we use the successive convex approximation (SCA) algorithm [27] to transform the non-convex optimization problems into a series of convex problems, thus obtaining the approximate solutions of the original problems. The optimization problem of delay can be written as

$$min\ T_v = F_v + G_v \quad (8)$$

where

$$F_v = (1-\beta-\gamma)\left(\frac{D_v}{\delta_{tr}^{dc}} + \frac{R_v}{\delta_{tr}^{cs}}\right) + \gamma\left(\frac{D_v}{\delta_{tr}^{ds}} + \frac{R_v}{\delta_{tr}^{ss}}\right) + \frac{R_v}{\delta_{tr}^{ds}} \quad (9)$$

and

$$G_v = max\left\{\frac{\beta D_v}{\eta_d}, \frac{(1-\beta-\gamma) D_v}{\eta_c}, T_{com}^{i}\right\} \quad (10)$$

For the task v and variable γ, F_v is further split into

$$F_v = f_1(\gamma) + f_2(\gamma) \tag{11}$$

$$f_1(\gamma) = \gamma\left(\frac{D_v}{\delta_{tr}^{ds}} + \frac{R_v}{\delta_{tr}^{ss}}\right) + \frac{R_v}{\delta_{tr}^{ds}} \tag{12}$$

$$f_2(\gamma) = (1-\beta-\gamma)\left(\frac{D_v}{\delta_{tr}^{dc}} + \frac{R_v}{\delta_{tr}^{cs}}\right) \tag{13}$$

where $f_2(\gamma)$ is linear with respect to $\gamma[n]$, n is the time interval (number of iterations). Then the problem (8) can be rewritten as

$$\widehat{\gamma}(\gamma[n]) = \arg\min_{\gamma\in[0,1]} T_v(\gamma) = F_v(\gamma) + G_v(\gamma) = f_1(\gamma) + f_2(\gamma) + G_v(\gamma) \tag{14}$$

Let $\widetilde{f}_1(\gamma;\gamma[n])$ approximately replaces $f_1(\gamma)$, then $f_1(\gamma)$ can be replaced by

$$\widetilde{f}_1(\gamma;\gamma[n]) = f_1(\gamma) + \nabla f_1(\gamma[n])^T(\gamma-\gamma[n]) + \frac{\tau_1}{2}\|\gamma-\gamma[n]\|^2 \tag{15}$$

then (14) can be approximated by

$$\widehat{\gamma}(\gamma[n]) = \arg\min_{\gamma\in[0,1]} \widetilde{f}_1(\gamma;\gamma[n]) + \pi_1(\gamma[n])^T(\gamma-\gamma[n]) + G_v(\gamma) \tag{16}$$

where $\pi_1(\gamma[n])$ is the gradient of $f_2(\gamma)$ with respect to $\gamma[n]$, be

$$\pi_1(\gamma[n]) = \nabla_\gamma f_2(\gamma[n]) = 2\left(\frac{1}{2}\sum_{p=1}^{2}\nabla f_p(\gamma[n])\right) - \nabla f_1(\gamma[n]) \tag{17}$$

Let $\widetilde{\pi}_1[n]$ approximately replaces $\pi_1(\gamma[n])$, where $\widetilde{\pi}_1[n] = 2\cdot y_1[n] - \nabla f_1(\gamma[n])$,

$$y_1[n+1] = w_{21}[n]y_1[n] + w_{12}[n]y_2[n] + (\nabla f_1(\gamma[n+1]) - \nabla f_1(\gamma[n])) \tag{18}$$

and $y_1[0] = \nabla f_1(\gamma[0])$.

In a similar way, we can figure out

$$\widehat{\beta}(\beta[n]) = \arg\min_{\beta\in[0,1]} \widetilde{f}_2(\beta;\beta[n]) + \pi_2(\beta[n])^T(\beta-\beta[n]) + G_v(\beta) \tag{19}$$

where $\pi_2(\beta[n])$ is the gradient of $f_1(\beta)$ with respect to $\beta[n]$.

As above, the non-convex problem (6) has been transformed into two convex optimization problems (16) and (19), which can then be solved by applying convex optimization methods. Similarly, an approximate solution to problem (7) can be obtained. Algorithm 1 shows the steps to further solve the delay and energy optimization problems. First, we start with initial arbitrarily values $[E_v]'_{v\in V}$, $[T_v]'_{v\in V}$ and stopping error ε_{error}, φ_{error} for the iterative algorithm. E_v' and T_v' start with small values, since the objective functions aim to reduce the system energy consumption and task execution delay as much as possible. In each iteration of Algorithm 1, we solve the convex optimization problem transformed from SCA algorithm using the gradient descent algorithm.

Algorithm 1: Resources Allocation in the SATCECN

Input: D_v, R_v, δ_{tr}^{dc}, δ_{tr}^{ds}, δ_{tr}^{cs}, δ_{tr}^{ss}, e_v^d, e_v^c, $e_v^s(i)$, e_{tr}^{dc}, $e_{tr}^{ds}(i)$, e_{tr}^{cs}, $e_{tr}^{ss}(i)$, e_{tr}^{sd}.
Initialize: ε_{error}, φ_{error}, E_v', T_v', $\forall v\in V$.

(Continued)

Algorithm 1 (continued)

While $\left(\left|T_v^{'} - T_v^{*}\right|_{v\epsilon V} \geq \varphi_{error}\right)$ **do**
  Solve problem (6) to get $[T_v]_{v\epsilon V}^{*}$;
  $T_v^{'} \leftarrow T_v^{*}$; $\forall v\epsilon V$
end
While $\left(\left|E_v^{'} - E_v^{*}\right|_{v\epsilon V} \geq \varepsilon_{error}\right)$ **do**
  Solve problem (7) to get $[E_v]_{v\epsilon V}^{*}$;
  $E_v^{'} \leftarrow E_v^{*}$; $\forall v\epsilon V$
end
Output: β, $\gamma_{i,i\epsilon\{1,2,\cdots,m\}}$, γ

5 Performance Evaluation

In this section, several experiments are done to evaluate the performance of the MCCC method in the SATCECN. Specifically, we compared the MCCC approach to the other three benchmark computing paradigms. In the first benchmark, the computation tasks are done entirely on the local terminal device, and this straightforward solution is called the *completely local computation*. The second benchmark is a *completely cloud computation* paradigm which sends all the computing tasks to the cloud for processing. The third benchmark is the *device and spatial cloud collaborative computation* paradigm. By comparing with the three benchmark computing paradigms, results show that the proposed MCCC approach can reduce both the system energy consumption and task completion delay.

To verify the proposed MCCC method, the experiments were run on an Intel(R) Core(TM) i7-6700HQ CPU with a speed of 2.6 GHz, 256 GB of storage, Windows 10 OS and Python with CVXPY (optimization tools). The computer simulations are performed in Python 3.11.2, where the considered parameters are listed in Table 1. The simulation experiments are performed with the aim of comparing the performance in terms of average task completion delay and weighted system energy consumption, which are defined as:

- Average task completion delay: The average time it takes for a task to complete data transmission, computation, and reception of computation results.
- Weighted system energy consumption: In each task calculation process, the weighted sum of the energy consumption generated by all terminal devices, satellite edge nodes, and spatial cloud nodes participating in the collaborative computing.

Table 1: Simulation parameters

Parameters	Value
Task size (D_v)	0–60 GB
Result size (R_v)	0–800 MB
δ_{tr}^{dc}, δ_{tr}^{ds}	822.5 Mbps
δ_{tr}^{cs}, δ_{tr}^{ss}	25 Mbps
The number of satellite user	20
Simulation scenario area	200 m × 200 m

(Continued)

Table 1 (continued)

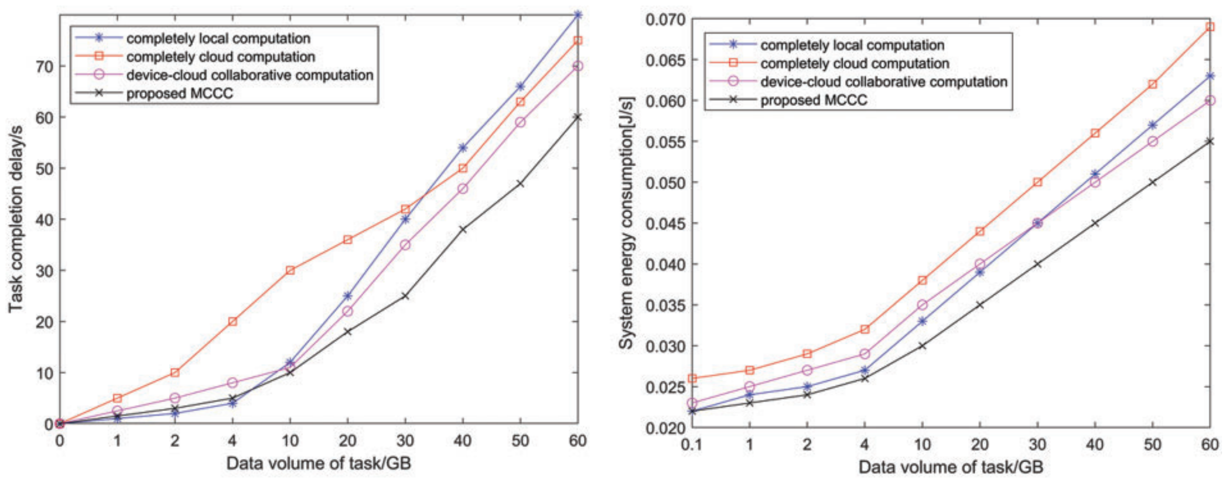
Parameters	Value
Channel bandwidth	250 MHz
Maximum number of spatial edge collaboration node	6
Maximum allowable execution delay	100 s

The experiments hypothesize that satellite users in the simulation scenario area randomly generate computing tasks of different sizes. The selection of datasets is essential for verifying and assessing the quality of any algorithm. A good dataset is necessary for providing reliable and effective outcomes. The simulation satellite network is established according to the dataset of satellites distribution in Starlink. The dataset includes LEO communication satellites deployed by the SpaceX Starlink project over North America, Asia, Europe, Oceania, and America. Each LEO constellation is connected to two MEO satellites and one GEO satellite. Satellites communicate with each other through inter-satellite laser links. There are usually frequent satellite communication task requests distributed within the coverage area of these satellites, which are very suitable for the cooperative processing in ISTN. In the MCCC-based solution, the algorithm runs 1,000 times for each computation task to eliminate accidental errors.

The first experiment compares the proposed MCCC solution against the other three benchmark computing paradigms in terms of average task completion delay and weighted system energy consumption. In order to investigate the advantages of the MCCC scheme itself, $\alpha = 0.5$ is set in this experiment. As shown in Fig. 3, the MCCC algorithm shows better results in terms of average task completion delay and weighted energy consumption of the system, when compared to the other three benchmark computing paradigms. For computing tasks with small amount of data, the user's local computing ability can basically meet the task requirements and complete the computing task with the fastest speed. However, as the amount of task data increases, the proposed MCCC method solves the limitation of insufficient local computing power, avoids the high propagation delay caused by cloud processing, and greatly shortens the completion time of tasks. Since edge cooperative computing nodes are located in low earth orbit, devices can transmit data with smaller transmission power, thereby reducing device energy consumption. The proposed MCCC algorithm allocates computing tasks to multiple cooperative nodes, not only alleviates the energy anxiety of a single satellite, but also saves the system energy overhead.

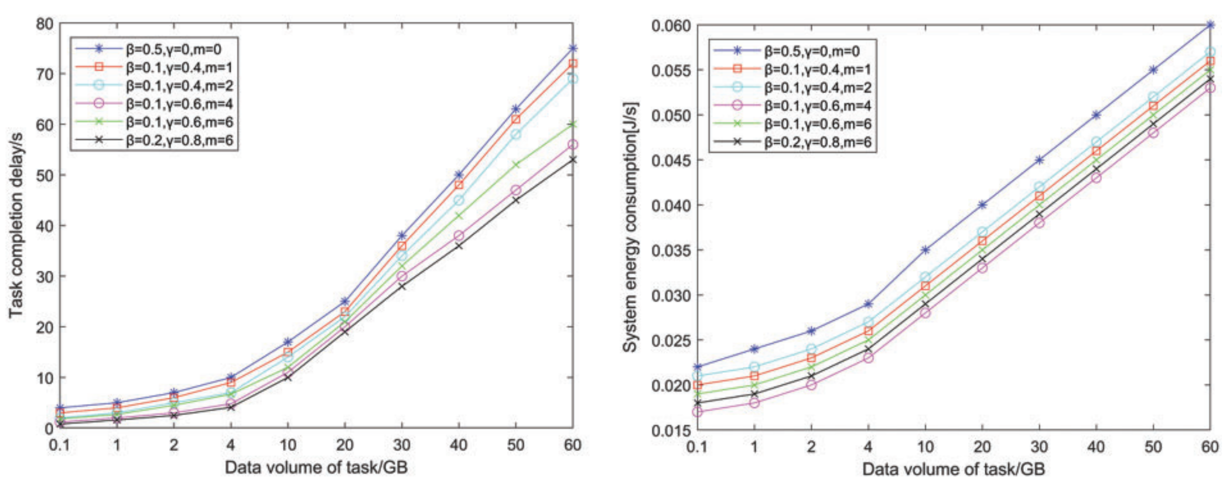
Then, experiments analyze the impact of the number of edge cooperative computing nodes and the allocation proportion of computing tasks among cooperative computing nodes on the average task completion delay and system weighted energy consumption. We write the number of satellite edge cooperative nodes that are involved in the task computation as m.

As shown in Fig. 4, when the number of spatial edge cooperative nodes available for task computation is limited, the average task completion delay mainly depends on the number of spatial edge cooperative nodes. The more edge collaborative nodes there are, the less tasks need to be transmitted to the cloud cooperative nodes, which can greatly reduce the average completion delay of tasks. When there are enough spatial edge cooperative nodes in the visible range of satellite users, the average task completion delay mainly depends on the allocation strategy of the computing tasks. At this time, the participation of spatial cloud cooperative nodes will lead to a longer task completion time due to the large propagation delay, so the computing tasks should be distributed between the device and multiple edge cooperative nodes. The weighted energy consumption of the system is affected by both the number of spatial edge cooperative computing nodes and the task allocation strategy. It will incur additional energy overhead to assign computationally intensive tasks to the spatial cloud computing center or a large number of spatial edge cooperative nodes. Evenly allocating computing tasks according to the computing power of device, spatial cloud nodes and each edge node can produce the optimal system energy consumption.



(a) (b)

Figure 3: Performance of MCCC and the other computing paradigms



(a) (b)

Figure 4: Performance of MCCC in different number of edge nodes and task allocation strategies

Since satellites, especially edge cooperative computing nodes located in low altitude orbit, are in an energy constrained state. Finally, the experiments analyze the influence of weight α on the weighted system energy consumption. In order to highlight the energy conditionality of spatial computing nodes, the three comparative experiments gradually increased the weight of the energy consumption of spatial cooperative nodes in the system energy consumption. Therefore, the weighted system energy consumption under different resource allocation strategies and the number of cooperative computing nodes is plotted when $\alpha = 0.5$, 0.4 and 0.3. As shown in Fig. 5, with the increase of the weight of the energy consumption of space cooperative nodes, the gap between the system energy consumption under different task allocation strategies is gradually widened. When α is smaller, the advantages of the proposed MCCC algorithm in saving system energy consumption are more prominent.

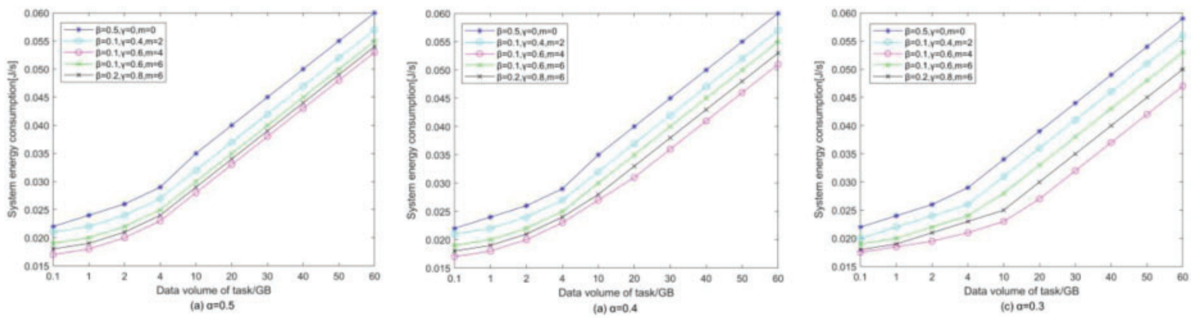


Figure 5: Performance of MCCC in system energy consumption under different α values

6 Conclusion and Future Work

Cloud computing and edge computing show strong processing capacity and fast response speed on the data side in the terrestrial core network, bringing huge performance improvement to the 4G and 5G networks. The introduction of cloud computing and edge computing resources into ISTN will be a key step in the development of the 6G network. In this paper, a new device-edge-cloud collaborative satellite-air-terrestrial layered architecture named SATCECN is studied, and its main application functions are discussed. Then this paper analyzes the roles of cloud computing and MEC in the SATCECN, and gives the possible deployment positions of cloud servers and edge servers in the SATCECN. On this basis, the process flow of typical tasks is described in detail. In view of reducing system energy consumption and task execution delay, this paper proposed a device-edge-cloud multi-node cross-layer collaborative computing method to solve the computing resources allocation problem. Comprehensive experiment results demonstrate that the outstanding performance of the MCCC method. Despite the above contributions, this paper leaves some deficiencies in the mobility of the satellite edge nodes in the SATCECN, the assignment strategy of computation tasks, the feedback mechanism of the task processing results, which will be our following research content. In addition, the proposed method will be combined with China Tiantong-1 satellite mobile communication system for case study and test in the future.

Acknowledgement: This work is supported by the Academic Discipline, Post-Graduate Education Project of the Beijing Municipal Commission of Education, and Fundamental Research Funds for the Central Universities, and the National Natural Science Foundation of China. This work is supported by the Institute of Intelligent Network and Information Security of Beijing Jiaotong University, which provides the experimental platform and technical support. We thank Feng Sun and Kaige Zhu for their guidance on the related theoretical knowledge.

Funding Statement: This work is supported by the Academic Discipline, Post-Graduate Education Project of the Beijing Municipal Commission of Education, and Fundamental Research Funds for the Central Universities under Grant 2022YJS015, and the National Natural Science Foundation of China under Grant 62173026.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. Mustafa, A. M. Buttar, G. S. Sajja, S. Gour and M. Naved, "Multitask learning for security and privacy in IoV (Internet of Vehicles)," *Autonomous Vehicles*, vol. 1, no. 12, pp. 217–233, 2022.

[2] R. Girau, "Coastal monitoring system based on social internet of things platform," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1260–1272, 2020.

[3] F. K. Shaikh, S. Karim, S. Zeadally and J. Nebhen, "Recent trends in internet-of-things-enabled sensor technologies for smart agriculture," *IEEE Internet of Things Journal*, vol. 9, no. 23, pp. 23583–23598, 2022.

[4] M. Mustafa, M. Alshare, D. Bhargava, R. Neware and B. Singh, "Perceived security risk based on moderating factors for blockchain technology applications in cloud storage to achieve secure healthcare systems," *Computational and Mathematical Methods in Medicine*, vol. 2022, pp. 611285, 2022.

[5] H. Wang, X. Xia, T. Song and Y. Xing, "Survey on space-air-ground integrated networks in 6G," in *2021 IEEE/CIC Int. Conf. on Communications in China (ICCC Workshops)*, Xiamen, China, vol. 2021, pp. 315–320, 2021.

[6] B. Smail, D. T. Sanchez, D. L. B. Peconcillo Jr., D. J. V. De Vera and D. A. D. Horteza, "Investigating different applications of internet of things towards identification of vulnerabilities, attacks and threats," *International Journal of Next-Generation Computing*, vol. 13, no. 3, 2022.

[7] J. A. Fraire, O. Iova and F. Valois, "Space-terrestrial integrated internet of things: Challenges and opportunities," *IEEE Communications Magazine*, vol. 60, no. 12, pp. 64–70, 2022.

[8] X. Zhu and C. Jiang, "Integrated satellite-terrestrial networks toward 6G: Architectures, applications, and challenges," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 437–461, 2022.

[9] X. Fang, W. Feng, T. Wei, Y. Chen and N. Ge, "5G embraces satellites for 6G ubiquitous IoT: Basic models for integrated satellite terrestrial networks," *IEEE Internet of Things Journal*, vol. 8, no. 18, pp. 14399–14417, 2021.

[10] H. Yao, L. Wang, X. Wang, Z. Lu and Y. Liu, "The space-terrestrial integrated network: An overview," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 178–185, 2018.

[11] L. Boero, R. Bruschi, F. Davoli, M. Marchese and F. Patrone, "Satellite networking integration in the 5G ecosystem: Research trends and open challenges," *IEEE Network*, vol. 32, no. 5, pp. 9–15, 2018.

[12] G. Giambene, S. Kota and P. Pillai, "Satellite-5G integration: A network perspective," *IEEE Network*, vol. 32, no. 5, pp. 25–31, 2018.

[13] Y. Bi, "Software defined space-terrestrial integrated networks: Architecture, challenges, and solutions," *IEEE Network*, vol. 33, no. 1, pp. 22–28, 2019.

[14] Y. Shi, Y. Cao, J. Liu and N. Kato, "A Cross-domain SDN architecture for multi-layered space-terrestrial integrated networks," *IEEE Network*, vol. 33, no. 1, pp. 29–35, 2019.

[15] X. Zhu, C. Jiang, L. Kuang, N. Ge and S. Guo, "Cooperative transmission in integrated terrestrial-satellite networks," *IEEE Network*, vol. 33, no. 3, pp. 204–210, 2019.

[16] W. Y. Liu, M. Peng, G. Shou, Y. Chen and S. Chen, "Toward edge intelligence: Multiaccess edge computing for 5G and internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6722–6747, 2020.

[17] J. Xia, P. Wang, B. Li and Z. Fei, "Intelligent task offloading and collaborative computation in multi-UAV-enabled mobile edge computing," *China Communications*, vol. 19, no. 4, pp. 244–256, 2022.

[18] N. Cheng, "Space/aerial-assisted computing offloading for IoT applications: A learning-based approach," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 5, pp. 1117–1129, 2019.

[19] Z. Zhang, W. Zhang and F. H. Tseng, "Satellite mobile edge computing: Improving QoS of high-speed satellite-terrestrial networks using edge computing techniques," *IEEE Network*, vol. 33, no. 1, pp. 70–76, 2019.

[20] C. Qiu, H. Yao, F. R. Yu, F. Xu and C. Zhao, "Deep Q-learning aided networking, caching, and computing resources allocation in software-defined satellite-terrestrial networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5871–5883, 2019.

[21] N. Zhang, S. Zhang, P. Yang, O. Alhussein and W. Zhuang, "Software defined space-air-ground integrated vehicular networks: Challenges and solutions," *IEEE Communications Magazine*, vol. 55, no. 7, pp. 101–109, 2017.

[22] L. Yan, M. W. Ahmad, M. Jawarneh, M. Shabaz and R. Raffik, "Single-input single-output system with multiple time delay PID control methods for UAV cluster multiagent systems," *Security and Communication Networks*, vol. 2022, pp. 3935143, 2022.

[23] S. Zhou, G. Wang, S. Zhang, Z. Niu and X. S. Shen, "Bidirectional mission offloading for agile space-air-ground integrated networks," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 38–45, 2019.

[24] W. Shi, "Multi-drone 3-D trajectory planning and scheduling in drone-assisted radio access networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 8145–8158, 2019.

[25] H. Zhang, L. L. Zhang, Y. H. Yan and L. L. Ci, "A review of gateway load balancing methods in connecting MANET into internet," in *2021 23rd Int. Conf. on Advanced Communication Technology (ICACT)*, PyeongChang, Korea (South), vol. 2021, pp. 330–335, 2021.

[26] P. K. Kollu, "Blockchain techniques for secure storage of data in cloud environment," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 11, pp. 1515–1522, 2021.

[27] A. Liu, V. K. N. Lau and B. Kananian, "Stochastic successive convex approximation for non-convex constrained stochastic optimization," *IEEE Transactions on Signal Processing*, vol. 67, no. 16, pp. 4189–4203, 2019.