Engineering 8 (2022) 96–107

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
6G Requirements, Vision, and Enabling Technologies—Article

MEC-Empowered Non-Terrestrial Network for 6G Wide-Area Time-Sensitive Internet of Things

Chengxiao Liu [a,b], Wei Feng [a,b,*], Xiaoming Tao [a,b], Ning Ge [a,b]

[a] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[b] Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 31 December 2020
Revised 11 June 2021
Accepted 29 August 2021
Available online 16 November 2021

Keywords:
Cell-free
Mobile edge computing
Non-terrestrial networks
Sixth-generation
Wide-area time-sensitive IoT

ABSTRACT

In the upcoming sixth-generation (6G) era, the demand for constructing a wide-area time-sensitive Internet of Things (IoT) continues to increase. As conventional cellular technologies are difficult to directly use for wide-area time-sensitive IoT, it is beneficial to use non-terrestrial infrastructures, including satellites and unmanned aerial vehicles (UAVs). Thus, we can build a non-terrestrial network (NTN) using a cell-free architecture. Driven by the time-sensitive requirements and uneven distribution of IoT devices, the NTN must be empowered using mobile edge computing (MEC) while providing oasis-oriented on-demand coverage for devices. Nevertheless, communication and MEC systems are coupled with each other under the influence of a complex propagation environment in the MEC-empowered NTN, which makes it difficult to coordinate the resources. In this study, we propose a process-oriented framework to design communication and MEC systems in a time-division manner. In this framework, large-scale channel state information (CSI) is used to characterize the complex propagation environment at an affordable cost, where a nonconvex latency minimization problem is formulated. Subsequently, the approximated problem is provided, and it can be decomposed into sub-problems. These sub-problems are then solved iteratively. The simulation results demonstrated the superiority of the proposed process-oriented scheme over other algorithms, implied that the payload deployments of UAVs should be appropriately predesigned to improve the efficiency of using resources, and confirmed that it is advantageous to integrate NTN with MEC for wide-area time-sensitive IoT.

1. Introduction

In future sixth-generation (6G) networks, the concentration of cutting-edge technologies will change from humans to intelligent machines [1]. In contrast to human beings, these machines are usually unevenly distributed in remote areas [1], which are built to accomplish time-sensitive tasks [2,3]. This scenario increases the demand for constructing a wide-area time-sensitive Internet of Things (IoT) in the upcoming 6G era [1–3].

However, terrestrial infrastructures are difficult to deploy in remote areas [4–6], indicating that terrestrial cellular networks have blind sides in terms of coverage ability [7]. Consequently, it is difficult to serve intelligent machines using conventional fourth-generation (4G) and fifth-generation (5G) technologies. Considering this challenge, it is beneficial to employ non-terrestrial infrastructures, including satellites and unmanned aerial vehicles (UAVs), for wide-area time-sensitive IoT. Thus, we can build a non-terrestrial network (NTN). In particular, an NTN is needed to provide oasis-oriented on-demand coverage for machines and accommodate the uneven distribution of machines; thus, the NTN should be designed under a cell-free architecture [8]. In addition, driven by the time-sensitive requirements of machines, data from machines must be processed by the NTN as quickly as possible. Therefore, satellite communications (SatCom)-on-the-move antennas and edge servers can be carried on UAVs to build high-speed links between satellites and UAVs [9], and rapidly process data with mobile edge computing (MEC) [10], respectively. Thus, an MEC-empowered NTN must be constructed using the cell-free architecture. Nevertheless, communications and MEC are coupled with each other in the NTN with a complex propagation environment that arises new challenges. First, owing to the complex propagation environment in the MEC-empowered NTN, realizing oasis-oriented on-demand coverage under a cell-free

* Corresponding author.
E-mail address: fengwei@tsinghua.edu.cn (W. Feng).

https://doi.org/10.1016/j.eng.2021.11.002

architecture is challenging [8]. Additionally, as communication and MEC systems are coupled with each other, simultaneous coordination of the resources is slightly complicated [10]. Hence, we investigate the design of an MEC-empowered NTN for wide-area time-sensitive IoT.

2. Literature review

For wide-area IoT, the narrow-band IoT (NB-IoT) is an enabling technique that was designed under conventional cellular architecture [11], whereas the long-range radio (LoRa) technique was proposed to further expand network coverage [12]. In addition, the design of time-sensitive networks (TSNs) has garnered extensive attention worldwide to serve time-sensitive machines, where industrial automation is a principal application scenario [13–15]. Lo Bello and Steiner [13] provided an overview of the applicability of TSNs to various industrial systems. Liang et al. [14] presented a comprehensive survey on wireless networks for the wireless industrial automation–factory automation (WIA-FA) technique and its applications. Luvisotto et al. [15] evaluated the feasibility of wireless high-performance (WirelessHP) technology for industrial wireless networks. These studies promoted the standardization of 5G ultra-reliable low-latency communication (URLLC) [16] and the industrial IoT [17] constructed by the Third Generation Partnership Project (3GPP).

Owing to the coverage holes of terrestrial cellular networks, NTN may become an advantageous technique for 6G networks, where the standardization of NTNs has been launched in 3GPP Release 16 [18]. In the future, the design of an NTN for supporting a wide-area time-sensitive IoT will be discussed in 3GPP Release 17 [19]. In the existing studies, satellite-enabled IoT has been widely discussed as it can provide ubiquitous coverage for wide-area IoT [20–22]. De Sanctis et al. [20] investigated the protocols and architectures for a satellite-based internet for remote objects. Cioni et al. [21] studied the opportunities and challenges of satellite-enabled massive machine-type communications (MMTC). Zhen et al. [22] proposed an optimal preamble design method that could adapt to the group-based random access pattern for satellite-based MMTC. However, satellite-enabled IoT systems undergo a high latency and low efficiency [20–22], which entangles meeting the requirements of intelligent machines [1].

In addition, UAVs have the potential to provide on-demand services for wide-area time-sensitive IoT [23–26]. In Ref. [23], a low-latency routing algorithm was proposed for UAV-enabled IoT, which was designed using a layered network architecture with a UAV swarm. The design of a UAV-enabled IoT-oriented network was proposed in Ref. [24] to support real-time remote virtual reality. In Ref. [25], the uplink (UL) power of IoT devices was optimized to design a UAV-assisted URLLC network. A UAV-assisted ubiquitous trust evaluation system was designed to reliably collect data from IoT devices [26]. To further improve the latency performance, UAV-enabled IoT was integrated with MEC [27–32]. In Ref. [27], the three-dimensional deployment of UAVs was optimized to support time-sensitive IoT, where UAVs were mounted as cloudlets. The average latency of users in UAV-aided MEC networks were minimized, as reported in Ref. [28]. In Ref. [29], the trajectories of UAVs were optimized for a smart IoT community, where an augmented reality-based use case was discussed. An energy-efficient multi-domain resource allocation scheme was proposed in Ref. [30] considering stringent latency requirements. In Ref. [31], an online UAV-mounted edge server dispatching scheme was proposed, where latency fairness among users was guaranteed with an efficient resource utilization. Additionally, a multi-UAV task offloading system was established that could transmit data from IoT devices to edge servers in a trustworthy manner [32]. Nevertheless, the UAV-enabled network usually lacks persistence and stability [33], which is an inevitable limitation for wide-area time-sensitive IoT.

Therefore, it is advantageous to jointly use satellites and UAVs with MEC for wide-area time-sensitive IoT [10,34,35]. In Ref. [34], Liu et al. presented a task-oriented intelligent architecture for IoT-oriented space–aerial–ground–aqua-integrated networks. Cheng et al. [10] investigated the concurrent design of computing resource allocation and task offloading strategies for IoT-oriented space–aerial–ground integrated networks, where stringent latency constraints were utilized and a learning-based approach was proposed. Cao et al. [35] discussed the coupling of trajectory design and task offloading strategies in an integrated satellite-UAV network under the influence of wind. Despite these achievements, when an NTN is integrated with MEC under a cell-free architecture, new challenges will be encountered. First, because of the complex propagation environment, NTN cannot perfectly acquire the channel state information (CSI), resulting in a complicated design of oasis-oriented on-demand coverage for machines under the cell-free architecture. Second, the resources cannot be readily coordinated in the MEC-empowered NTN because communication and MEC systems are coupled with each other. In our previous study [8], we discussed the cell-free coverage patterns of integrated satellite-UAV networks. In this study, we advance the investigation to the design of an MEC-empowered NTN for wide-area time-sensitive IoT. The relationships between the existing technologies and certain research areas are summarized in Table 1 [11,12,14–17,20–22].

3. Study contributions

In this study, we investigated the design of an MEC-empowered NTN for a wide-area time-sensitive IoT. In particular, we focused on the design of NTN, which consists of hierarchically integrated satellites and UAVs considering the overall communication and computing latency as the metric of latency performance. The MEC-empowered NTN is designed under a process-oriented framework in a time-division manner [8] to satisfy the service requirements of wide-area time-sensitive IoT, where a latency minimization problem is formulated using a large-scale CSI. Subsequently, a process-oriented joint resource orchestration scheme is proposed to solve the latency minimization problem. The main contributions of this study are summarized as follows:

(1) A process-oriented framework is presented for an MEC-empowered NTN. This framework can jointly design communication and MEC systems in a time-division manner for hierarchically integrated satellites and UAVs. Subsequently, an overall communication and computing latency minimization problem is formulated, where large-scale CSI is used to characterize complex propagation environments at an affordable cost.

(2) As the latency minimization problem is a nonconvex stochastic optimization problem, we first prove that the original problem can be transformed into a simplified form. Subsequently, we propose an approximation of the simplified problem, which can be further decomposed into sub-problems according to the properties of the overall communication and computing efficiency function.

(3) We propose a joint power allocation and data stream scheduling scheme to solve sub-problems, where block coordinate descent and successive convex approximation techniques are applied. The process-oriented joint resource orchestration scheme is derived iteratively.

The remainder of this paper is organized as follows. We introduce the system model and the process-oriented framework in Section 4. In Section 5, the solution of the latency minimization

Table 1
Existing technologies and our concentrations.

Specific area	Latency requirement	Coverage pattern	Technology	Refs.
Wide-area IoT	Large latency allowed	Cellular-based	NB-IoT	[11]
		Expanded cellular	LoRa	[12]
		Ubiquitous	Satellite-based IoT	[20–22]
TSN	Sensitive to latency	Indoor	WIA-FA	[14]
			WirelessHP	[15]
		Cellular-based	3GPP 5G URLLC	[16]
			3GPP Industrial IoT	[17]
Wide-area time-sensitive IoT		Oasis-oriented under a cell-free architecture	MEC-empowered NTN	

problem is presented, where a joint power allocation and data stream scheduling scheme is introduced. The simulation results and discussions are presented in Section 6, and the conclusions are drawn in Section 7.

4. System model

Fig. 1 illustrates an MEC-empowered NTN with hierarchically integrated satellite and UAVs, which has U single-antenna IoT devices, K UAVs in a swarm that are equipped with an MEC server and M antennas, and a satellite that can transmit data back to the cloud server via a gateway. We assume that the UAVs fly around the IoT devices following a predetermined circular trajectory. This mode can save energy with guaranteed stability [36]. To accommodate the distribution of devices in wide-area time-sensitive IoT, a hierarchical NTN is designed under a cell-free architecture [8], where an oasis-oriented coverage pattern can be observed. Based on such coverage patterns, the associations between the devices and UAVs can be predetermined. For simplifying the mathematical analysis, the indicator set of user association is denoted as $z = \{z_{u,k}\}$, where $z_{u,k} = 1$ means that the uth device is associated with the kth UAV.

In practical systems, the computing ability of each IoT device is usually weak; thus, devices must upload data to the satellite or UAVs to accomplish computation-intensive yet time-sensitive tasks [10]. After the cloud server successfully receives all the data from devices, the entire process of communication and computing is completed [37,38]. We assume that the uth device has D_u data to be uploaded. The communication and computing process is designed under a process-oriented framework to manage the influence of UAV movement on data transmission, which can reduce the complexity of optimizing the entire process [8]. As illustrated in Fig. 2, the entire process is divided into N_T segmentations. The parameters of the MEC-empowered NTN are updated at the beginning of the segmentation, and are assumed to be constant during each segmentation and possibly vary with each other in different segmentations. The update interval of system parameters is denoted as δ_T, and the overall communication and computing latency can be expressed as $T_\mathrm{total} = N_\mathrm{T}\delta_\mathrm{T} + \epsilon_\mathrm{a}$, where ϵ_a is the total propagation time of the electromagnetic wave. In particular, in the tth segmentation, the uth device can send a ratio of η_t^L data directly to the satellite, the ratio of $\eta_{u,t}^\mathrm{S}$ data to the satellite via device–UAV and UAV–satellite links, and the ratio of $\eta_{u,t}^\mathrm{C}$ data via device–UAV links to on-board MEC servers for computing. Thus, we have Eqs. (1) and (2) as the practical constraints for these ratios. After the data are computed by the MEC servers, the computational results are transmitted from the MEC servers to the satellite via UAV-satellite links. For simplifying the mathematical analysis, we assume that the output data size is proportional to that of the input data for MEC servers [39,40], where the proportion of the data from the uth device is denoted as ζ_u.

$$\eta_t^\mathrm{L} + \eta_{u,t}^\mathrm{S} + \eta_{u,t}^\mathrm{C} = 1 \ \ \forall u, t \tag{1}$$

$$0 \le \eta_t^\mathrm{L},\ \eta_{u,t}^\mathrm{S},\ \eta_{u,t}^\mathrm{C} \ \le 1 \ \ \forall u, t \tag{2}$$

In each segmentation of the process, data from IoT devices are first transmitted to the satellite or UAVs. Under the cell-free architecture, all devices are assumed to share the same frequency band

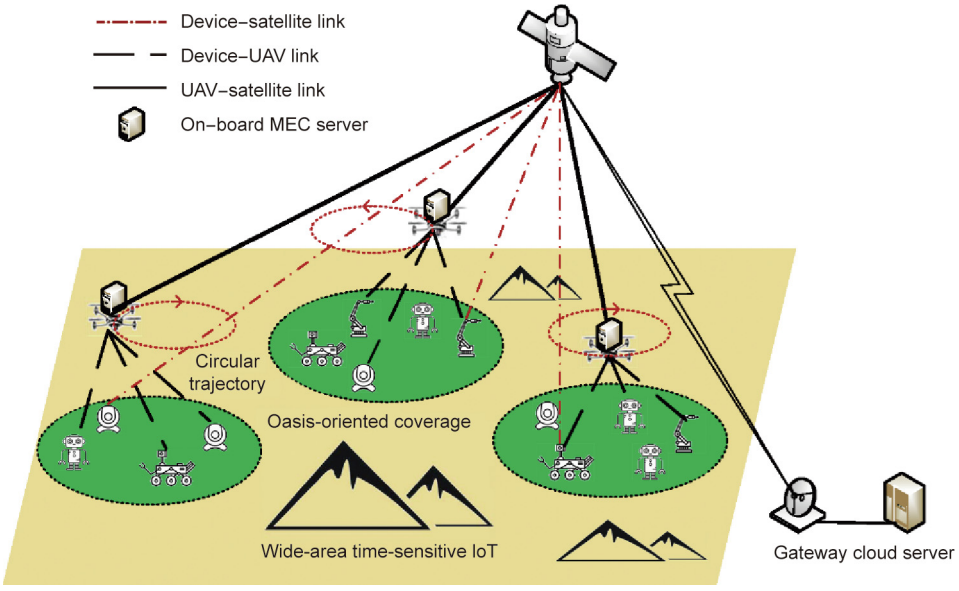

Fig. 1. Illustration of an MEC-empowered hierarchical NTN for wide-area time-sensitive IoT.

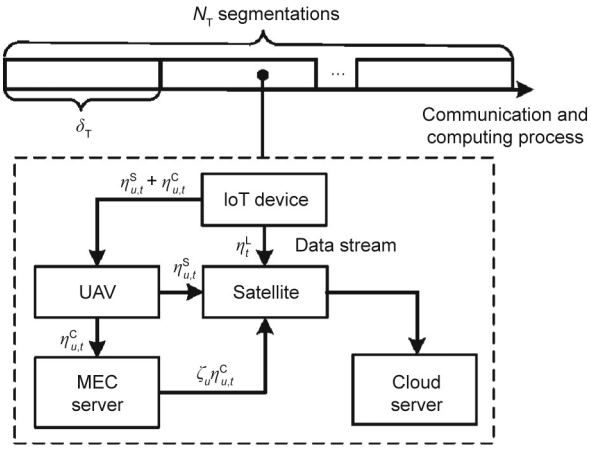

Fig. 2. Diagram of the process-oriented framework in the MEC-empowered NTN.

[8], where the bandwidth is denoted as B. When an IoT device is directly connected to the satellite, we assume that the UL rate between the device and satellite is constant [41], which is R^{L}. Moreover, when IoT devices are connected with UAVs, they consist of a multiuser multiple-input-multiple-output (MU-MIMO) UL system for data transmission. Therefore, the receive symbol of the uth user from the kth UAV in the tth segmentation is formulated as follows:

$$\boldsymbol{y}_{u,k,t} = \boldsymbol{h}_{u,k,t} x_{u,t} + \sum_{v=1, v\neq u}^{U} \boldsymbol{h}_{v,k,t} x_{v,t} + \boldsymbol{n}_{u,k,t} \tag{3}$$

where $x_{u,t}$ and $x_{v,t}$ are the transmitted symbol, $\boldsymbol{n}_{u,k,t}$ is the additive white Gaussian noise that satisfies $\boldsymbol{n}_{u,k,t} \sim N(0, \sigma^2 \boldsymbol{I}_M)$, σ^2 is the variance of noise, $\boldsymbol{I}_M$ denotes the M-dimensional identity matrix and $\boldsymbol{h}_{u,k,t}$ is the channel vector, which is written as

$$\boldsymbol{h}_{u,k,t} = s_{u,k,t} l_{u,k} \boldsymbol{a}_{u,k} \tag{4}$$

where $s_{u,k,t}$ denotes fast-varying small-scale parameters with identical distributions whose phases are uniformly distributed in $[0, 2\pi]$, whereas their amplitudes follow the Nakagami-m distribution and the probability density function is [36]

$$f_{|s|}(z) = \frac{2m^m}{\Gamma(m)\Omega^m} z^{2m-1} \mathrm{e}^{-\frac{mz^2}{\Omega}} \tag{5}$$

where m and Ω are parameters of the Nakagami-m distribution, and $\Gamma(m)$ denotes the gamma function with respect to m. These small-scale parameters are assumed to be independent of each other for different values of u, k, and t [42]. Additionally, $l_{u,k}$ is the slowly varying path loss of the UAV channel, which is expressed as follows [36]:

$$l_{u,k} = 10^{-\frac{L_{u,k}}{20}} \tag{6}$$

$$L_{u,k} = \frac{A_0}{1 + a\mathrm{e}^{-b\left(\theta_{u,k} - a\right)}} + B_0 \tag{7}$$

where $A_0 = \eta_{\mathrm{LOS}} - \eta_{\mathrm{NLOS}}$; $B_0 = 20\lg\left(d_{u,k}\right) + 20\lg\left(\frac{4\pi f}{c}\right) + \eta_{\mathrm{NLOS}}$; f is the carrier frequency; c denotes the speed of light; η_{LOS}, η_{NLOS}, a, and b are constants related to the propagation environment; $d_{u,k}$ is the distance between the device and UAV; $\theta_{u,k} = \frac{180}{\pi} \arcsin\left(\frac{h_k}{d_{u,k}}\right)$ represents the azimuth angle of the device-UAV link; h_k is the height of the kth UAV; and $\boldsymbol{a}_{u,k} \in \mathbb{C}^{M\times 1}$ is the array manifold vector of the receiver antenna array. We assume that uniform linear arrays (ULAs) can be carried on UAVs; thus, we have the following expressions [43]:

$$\boldsymbol{a}_{u,k} = \left[1, \mathrm{e}^{j\frac{2\pi f d_0}{c}\cos\left(\theta_{u,k}\right)}, \ldots, \mathrm{e}^{j\frac{2\pi f d_0}{c}(M-1)\cos\left(\theta_{u,k}\right)}\right]^{\mathrm{T}} \tag{8}$$

where T is the transpose symbol; j is the imaginary unit; and d_0 is the distance between adjacent antennas.

In this system, the process of data transmission from devices to UAVs is designed prior to UAV takeoff, which its time scale is considerably larger than the channel coherence time. Therefore, using pilot symbols, the UL CSI in Eq. (4) can be accurately estimated by UAVs within the channel coherence time; however, such a CSI cannot be perfectly acquired prior to UAV takeoff considering the large time scale of the entire process. Consequently, a perfect CSI is difficult to use when designing a data transmission process. In particular, we regard the position-related parameters, that is, $l_{u,k}$ and $\boldsymbol{a}_{u,k}$ as slowly varying large-scale channel parameters, which can be perfectly acquired using radio maps in practical systems [5]. These parameters are assumed to be constant throughout the entire process. In contrast, $s_{u,k,t}$ varies rapidly owing to the movement of UAVs, and only its distribution is known. Under these assumptions, the efficiency of data transmission in each segmentation of the process can be evaluated using the ergodic rate [8]. In addition, it is reasonable to assume that minimum mean square error (MMSE) detection is used at the receiver [37], where the detection vector for the uth device at the kth UAV in the tth segmentation is denoted as $\boldsymbol{w}_{u,k,t}$ [44]. Consequently, the UL ergodic rate of the uth device at the kth UAV in the tth segmentation can be formulated as follows [45]:

$$R^{\mathrm{UL}}_{u,k,t}(\boldsymbol{P}) = (1-\gamma_{\mathrm{UL}}) B \mathbf{E}\left\{\log_2\left(1 + \frac{p_{u,t}\left|\boldsymbol{w}^{\mathrm{H}}_{u,k,t}\boldsymbol{h}_{u,k,t}\right|^2}{\sum_{v=1,v\neq u}^{U} p_{v,t}\left|\boldsymbol{w}^{\mathrm{H}}_{u,k,t}\boldsymbol{h}_{v,k,t}\right|^2 + \left|\left|\boldsymbol{w}_{u,k,t}\right|\right|^2 \sigma^2}\right)\right\} \tag{9}$$

where $\mathbf{E}$ denotes the symbol of mathematical expectation, H is the conjugate transpose symbol, γ_{UL} denotes the fraction of transmitted signals that are used as reference signals; hence, $R^{\mathrm{UL}}_{u,k,t}(\boldsymbol{P})$ can be regarded as an achievable net rate [46], $\boldsymbol{P} = \left(p_{u,t}\right) \in \mathbb{R}^{U\times N_{\mathrm{T}}}$ denotes the power matrix, and $p_{u,t} = \mathbf{E}\left\{x^{\mathrm{H}}_{u,t} x_{u,t}\right\}$ is the signal power.

After UAVs receive data streams from devices, these streams are further scheduled for communication and computing. To guarantee the stability, this system is assumed to work in a nonblocking mode [39], where any packet of data in the data stream can be transmitted from devices to the cloud server without any waiting time. Therefore, the constraints of the data streams are derived as follows [39]:

$$\sum_{u=1}^{U} \eta^{\mathrm{C}}_{u,t} z_{u,k} R^{\mathrm{UL}}_{u,k,t}(\boldsymbol{P}) \leq R^{\mathrm{C}}_k, \quad \forall k, t \tag{10}$$

$$\sum_{u=1}^{U} \left(\eta^{\mathrm{S}}_{u,t} + \zeta_u \eta^{\mathrm{C}}_{u,t}\right) z_{u,k} R^{\mathrm{UL}}_{u,k,t}(\boldsymbol{P}) \leq R^{\mathrm{S}}_k, \quad \forall k, t \tag{11}$$

where R^{C}_k is the average throughput of the MEC server, R^{S}_k represents the data rate of the UAV–satellite link, and both are parameters of the kth UAV. Furthermore, for the uth device in the tth segmentation, the average overall communication and computing efficiency are expressed as follows:

If only a satellite is used for communication, we have

$$R^{\mathrm{a}}_{u,t} = R^{\mathrm{L}} \tag{12a}$$

If the MEC server on the UAV is not used for computing, we have

$$R^{\mathrm{a}}_{u,t} = \left(\frac{\eta^{\mathrm{L}}_t}{R^{\mathrm{L}}} + \frac{\eta^{\mathrm{S}}_{u,t}}{\sum_{k=1}^{K} z_{u,k} R^{\mathrm{UL}}_{u,k,t}(\boldsymbol{P})} + \frac{\eta^{\mathrm{S}}_{u,t}}{\sum_{k=1}^{K} z_{u,k} R^{\mathrm{S}}_k}\right)^{-1} \tag{12b}$$

If the MEC server on the UAV is used for computing, we have

$$R_{u,t}^{\mathrm{a}}=\left(\frac{\eta_t^{\mathrm{L}}}{R^{\mathrm{L}}}+\frac{\eta_{u,t}^{\mathrm{S}}+\eta_{u,t}^{\mathrm{C}}}{\sum_{k=1}^{K}z_{u,k}R_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})}+\frac{\eta_{u,t}^{\mathrm{S}}+\zeta_u\eta_{u,t}^{\mathrm{C}}}{\sum_{k=1}^{K}z_{u,k}R_k^{\mathrm{S}}}+\frac{\eta_{u,t}^{\mathrm{C}}}{\sum_{k=1}^{K}z_{u,k}R_k^{\mathrm{C}}}\right)^{-1} \tag{12c}$$

As shown in Eqs. (12a)–(12c), $R_{u,t}^{\mathrm{a}}$ has different values in various cases. The reason is that data are transmitted from devices to the satellite and UAVs in sequential packets in practical systems. The update interval should be longer than the packet transmission time to maintain the system stable. Fig. 2 shows that at least a packet transmission time of $2\epsilon_0$ is required to send data to the satellite via device–UAV links and UAV–satellite links (ϵ_0 denotes the transmission time of data packet), whereas at least a packet transmission time of $3\epsilon_0$ is consumed to send data to the satellite via device–UAV links, MEC servers, and UAV–satellite links; where we assume that ε_0 denotes the transmission time of data packet. Thus, UAVs and MEC servers can only be used for communication and computing when the update interval is sufficiently large, which results different values for $R_{u,t}^{\mathrm{a}}$.

Based on Eqs. (1)–(12), the latency minimization problem can be formulated as follows:

$$\min_{\boldsymbol{P},\eta,\delta_{\mathrm{T}}} N_{\mathrm{T}}\delta_{\mathrm{T}} \tag{13a}$$

$$\text{s.t.}\ \sum_{t=1}^{N_{\mathrm{T}}}R_{u,t}^{\mathrm{a}}\delta_{\mathrm{T}}\geq D_u,\ \ \forall u \tag{13b}$$

$$\sum_{u=1}^{U}\eta_{u,t}^{\mathrm{C}}z_{u,k}R_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})\leq R_k^{\mathrm{C}},\ \ \forall k,t \tag{13c}$$

$$\sum_{u=1}^{U}\left(\eta_{u,t}^{\mathrm{S}}+\zeta_u\eta_{u,t}^{\mathrm{C}}\right)z_{u,k}R_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})\leq R_k^{\mathrm{S}},\ \ \forall k,t \tag{13d}$$

$$0\leq p_{u,t}\leq P_{\max},\ \ \forall u,t \tag{13e}$$

$$\eta_t^{\mathrm{L}}+\eta_{u,t}^{\mathrm{S}}+\eta_{u,t}^{\mathrm{C}}=1,\ \ \forall u,t \tag{13f}$$

$$0\leq\eta_t^{\mathrm{L}},\ \eta_{u,t}^{\mathrm{S}},\ \eta_{u,t}^{\mathrm{C}}\leq 1,\ \ \forall u,t \tag{13g}$$

where $\eta=\left\{\eta_t^{\mathrm{L}},\eta_{u,t}^{\mathrm{S}},\eta_{u,t}^{\mathrm{C}}\right\}$ $(\forall u,t)$ denotes the set of variables, N_{T} is predetermined and δ_{T} is optimized to find the minimum overall latency; Eq. (13b) guarantees that the service requirements of all devices are satisfied; Eqs. (13c) and (13d) are rate constraints of data streams; Eq. (13e) denotes the power constraints of the devices, where $P_{\max}$ is the maximum transmission power; Eq. (13f) represents the practical constraints of data stream scheduling; and Eq. (13g) shows the range of the variables.

5. Joint power allocation and data stream scheduling under the process-oriented framework

5.1. Problem decomposition

Observing Eqs. (13a)–(13d) reveal that Eq. (13) is a nonconvex stochastic optimization problem, which is difficult to solve directly using the existing tools. An approximation of $R_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})$ is provided as follows to further simplify Eq. (13) [45,47]:

$$\widehat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})=(1-\gamma_{\mathrm{UL}})B\log_2\left(1+\frac{p_{u,t}\theta_{u,u,k}}{\sum_{v=1,v\neq u}^{U}p_{v,t}\theta_{u,v,k}+\sigma^2}\right) \tag{14}$$

where

$$\theta_{u,v,k}=\mathbf{E}\left\{\frac{\left|\boldsymbol{w}_{u,k,t}^{\mathrm{H}}\boldsymbol{h}_{v,k,t}\right|^2}{\left\|\boldsymbol{w}_{u,k,t}\right\|^2}\right\} \tag{15}$$

According to Eqs. (14) and (15), $\boldsymbol{\theta}=\left\{\theta_{u,v,k}\right\}$ $(\forall u,v,k)$ is calculated prior to resource orchestration, and the parameters in $\boldsymbol{\theta}$ can be regarded as deterministic parameters, indicating that the expectation operator in Eq. (14) can be eliminated. Fig. 3 demonstrates that the proposed approximation is accurate with randomly generated channel vectors and transmission power, whereas such an approximation can also be justified on the basis of the central limit theorem and Jensen inequalities [48].

Based on Eqs. (14) and (15), $R_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})$ in Eqs. (13b)–(13d) can be replaced by $\widehat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})$. Subsequently, the problem in Eq. (13) is further decomposed into three sub-problems with respect to the segmentation of $R_{u,t}^{\mathrm{a}}$ that are illustrated as follows:

$$\min_{\boldsymbol{P},\eta,\delta_{\mathrm{T}}} \delta_{\mathrm{T}} \tag{16a}$$

$$\text{s.t.}\ \ N_{\mathrm{T}}\delta_{\mathrm{T}}R^{\mathrm{L}}\geq D_u,\ \ \forall u \tag{16b}$$

$$\sum_{u=1}^{U}\eta_{u,t}^{\mathrm{C}}z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})\leq R_k^{\mathrm{C}},\ \ \forall k,t \tag{16c}$$

$$\sum_{u=1}^{U}\left(\eta_{u,t}^{\mathrm{S}}+\zeta_u\eta_{u,t}^{\mathrm{C}}\right)z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})\leq R_k^{\mathrm{S}},\ \ \forall k,t \tag{16d}$$

$$0\leq p_{u,t}\leq P_{\max},\ \ \forall u,t \tag{16e}$$

$$\eta_t^{\mathrm{L}}+\eta_{u,t}^{\mathrm{S}}+\eta_{u,t}^{\mathrm{C}}=1,\ \ \forall u,t \tag{16f}$$

$$0\leq\eta_t^{\mathrm{L}},\ \eta_{u,t}^{\mathrm{S}},\ \eta_{u,t}^{\mathrm{C}}\ \leq 1,\ \ \forall u,t \tag{16g}$$

$$\delta_{\mathrm{T}}\geq\epsilon_0 \tag{16h}$$

$$\min_{\boldsymbol{P},\eta,\delta_{\mathrm{T}}} \delta_{\mathrm{T}} \tag{17a}$$

$$\text{s.t.}\ \sum_{t=1}^{N_{\mathrm{T}}}\delta_{\mathrm{T}}\left(\frac{\eta_t^{\mathrm{L}}}{R^{\mathrm{L}}}+\frac{\eta_{u,t}^{\mathrm{S}}+\eta_{u,t}^{\mathrm{C}}}{\sum_{k=1}^{K}z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})}+\frac{\eta_{u,t}^{\mathrm{S}}+\zeta_u\eta_{u,t}^{\mathrm{C}}}{\sum_{k=1}^{K}z_{u,k}R_k^{\mathrm{S}}}+\frac{\eta_{u,t}^{\mathrm{C}}}{\sum_{k=1}^{K}z_{u,k}R_k^{\mathrm{C}}}\right)^{-1}\geq D_u,\ \ \forall u \tag{17b}$$

$$\sum_{u=1}^{U}\eta_{u,t}^{\mathrm{C}}z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})\leq R_k^{\mathrm{C}},\ \ \forall k,t \tag{17c}$$

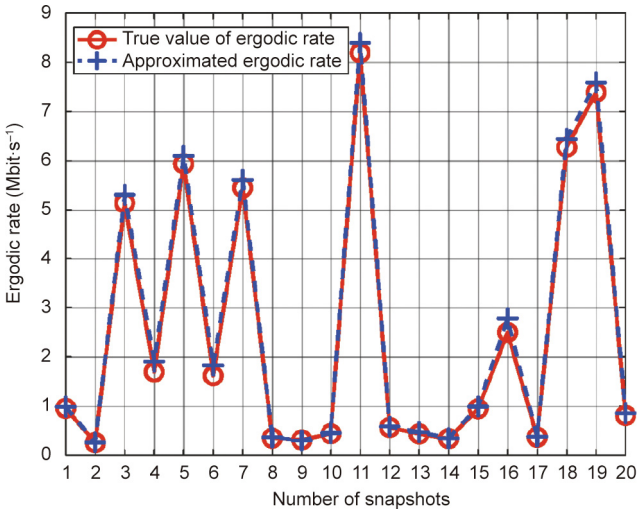

Fig. 3. Numerical evaluations of the approximated ergodic rate accuracy.

$$\sum_{u=1}^{U}\left(\eta_{u,t}^{\mathrm{S}}+\zeta_u\eta_{u,t}^{\mathrm{C}}\right)z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})\le R_k^{\mathrm{S}},\ \ \forall k,t \tag{17d}$$

$$0\le p_{u,t}\le P_{\max},\ \ \forall u,t \tag{17e}$$

$$\eta_t^{\mathrm{L}}+\eta_{u,t}^{\mathrm{S}}+\eta_{u,t}^{\mathrm{C}}=1,\ \ \forall u,t \tag{17f}$$

$$0\le\eta_t^{\mathrm{L}},\ \eta_{u,t}^{\mathrm{S}},\ \eta_{u,t}^{\mathrm{C}}\le 1,\ \ \forall u,t \tag{17g}$$

$$\delta_{\mathrm{T}}\ge 3\epsilon_0 \tag{17h}$$

$$\min_{\boldsymbol{P},\eta,\delta_{\mathrm{T}}}\ \delta_{\mathrm{T}} \tag{18a}$$

$$\text{s.t.}\ \sum_{t=1}^{N_{\mathrm{T}}}\delta_{\mathrm{T}}\left(\frac{\eta_t^{\mathrm{L}}}{R^{\mathrm{L}}}+\frac{\eta_{u,t}^{\mathrm{S}}}{\sum_{k=1}^{K}z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})}+\frac{\eta_{u,t}^{\mathrm{S}}}{\sum_{k=1}^{K}z_{u,k}R_k^{\mathrm{S}}}\right)^{-1}\ge D_u,\ \ \forall u \tag{18b}$$

$$\sum_{u=1}^{U}\eta_{u,t}^{\mathrm{C}}z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})\le R_k^{\mathrm{C}},\ \ \forall k,t \tag{18c}$$

$$\sum_{u=1}^{U}\left(\eta_{u,t}^{\mathrm{S}}+\zeta_u\eta_{u,t}^{\mathrm{C}}\right)z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})\le R_k^{\mathrm{S}},\ \ \forall k,t \tag{18d}$$

$$0\le p_{u,t}\le P_{\max},\ \ \forall u,t \tag{18e}$$

$$\eta_t^{\mathrm{L}}+\eta_{u,t}^{\mathrm{S}}+\eta_{u,t}^{\mathrm{C}}=1,\ \ \forall u,t \tag{18f}$$

$$0\le\eta_t^{\mathrm{L}},\ \eta_{u,t}^{\mathrm{S}},\ \eta_{u,t}^{\mathrm{C}}\le 1,\ \ \forall u,t \tag{18g}$$

$$\delta_{\mathrm{T}}\ge 2\epsilon_0 \tag{18h}$$

It is not difficult to certify that the sub-problems in Eqs. (16)–(18) are independent of each other. For simplicity of notation, if we have $(\boldsymbol{P},\eta)$ as the solution to Eqs. (16), (17), or (18), the corresponding objective function is expressed as $\delta_{\mathrm{T}}(\boldsymbol{P},\eta)$. Subsequently, we can discuss the solutions to these sub-problems individually.

5.2. Solution to Eq. (16)

According to the constraints in Eq. (16), we find that Eqs. (16c)–(16g) have no influence on the objective function in Eq. (16a). Therefore, Eq. (16) can be equivalently transformed into

$$\min_{\boldsymbol{P},\eta,\delta_{\mathrm{T}}}\ \delta_{\mathrm{T}} \tag{19a}$$

$$\text{s.t.}\ N_{\mathrm{T}}\delta_{\mathrm{T}}R^{\mathrm{L}}\ge D_u,\ \ \forall u \tag{19b}$$

$$\delta_{\mathrm{T}}\ge\epsilon_0 \tag{19c}$$

where the solution can be directly expressed as

$$\eta_t^{\mathrm{L}}=1,\ \eta_{u,t}^{\mathrm{S}}=\eta_{u,t}^{\mathrm{C}}=0,\ p_{u,t}=P_{\max} \tag{20}$$

The minimum update interval of system parameters becomes

$$\delta_{\mathrm{T}}=\max\left\{\epsilon_0,\frac{D_u}{N_{\mathrm{T}}R^{\mathrm{L}}}\right\},\ \ \forall u \tag{21}$$

Remark 1: The solution to Eq. (16) provides the joint resource orchestration scheme when we use only the satellite to transmit data. Intuitively, we show that the minimum overall latency is achieved by this strategy when D_u is sufficiently small. This is because the overall latency may be less than $2\epsilon_0$ if we only use the satellite for data transmission; however, the latency is at least $2\epsilon_0$ if UAVs are used, as shown in Eqs. (16h), (17h), and (18h). This intuition can be further verified by the simulation results.

5.3. Solution to Eq. (17)

Owing to the coupling of $\boldsymbol{P}$ and η in Eqs. (17b)–(17d), Eq. (17) is nonconvex and difficult to solve directly. To solve this problem, we use the block coordinate descent technique to decompose Eq. (17) into two sub-problems [8], which are formulated as

$$\min_{\boldsymbol{P}^i,\delta_{\mathrm{T}}}\ \delta_{\mathrm{T}} \tag{22a}$$

$$\text{s.t.}\ \sum_{t=1}^{N_{\mathrm{T}}}\delta_{\mathrm{T}}\left(\frac{\eta_t^{\mathrm{L},i-1}}{R^{\mathrm{L}}}+\frac{\eta_{u,t}^{\mathrm{S},i-1}+\eta_{u,t}^{\mathrm{C},i-1}}{\sum_{k=1}^{K}z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}\left(\boldsymbol{P}^i\right)}+\frac{\eta_{u,t}^{\mathrm{S},i-1}+\zeta_u\eta_{u,t}^{\mathrm{C},i-1}}{\sum_{k=1}^{K}z_{u,k}R_k^{\mathrm{S}}}+\frac{\eta_{u,t}^{\mathrm{C},i-1}}{\sum_{k=1}^{K}z_{u,k}R_k^{\mathrm{C}}}\right)^{-1}\ge D_u,\ \ \forall u \tag{22b}$$

$$\sum_{u=1}^{U}\eta_{u,t}^{\mathrm{C},i-1}z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}\left(\boldsymbol{P}^i\right)\le R_k^{\mathrm{C}},\ \ \forall k,t \tag{22c}$$

$$\sum_{u=1}^{U}\left(\eta_{u,t}^{\mathrm{S},i-1}+\zeta_u\eta_{u,t}^{\mathrm{C},i-1}\right)z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}\left(\boldsymbol{P}^i\right)\le R_k^{\mathrm{S}},\ \ \forall k,t \tag{22d}$$

$$0\le p_{u,t}^i\le P_{\max},\ \ \forall u,t \tag{22e}$$

$$\delta_{\mathrm{T}}\ge 3\epsilon_0 \tag{22f}$$

$$\min_{\eta^i,\delta_{\mathrm{T}}}\ \delta_{\mathrm{T}} \tag{23a}$$

$$\text{s.t.}\ \sum_{t=1}^{N_{\mathrm{T}}}\delta_{\mathrm{T}}\left(\frac{\eta_t^{\mathrm{L},i}}{R^{\mathrm{L}}}+\frac{\eta_{u,t}^{\mathrm{S},i}+\eta_{u,t}^{\mathrm{C},i}}{\sum_{k=1}^{K}z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}\left(\boldsymbol{P}^i\right)}+\frac{\eta_{u,t}^{\mathrm{S},i}+\zeta_u\eta_{u,t}^{\mathrm{C},i}}{\sum_{k=1}^{K}z_{u,k}R_k^{\mathrm{S}}}+\frac{\eta_{u,t}^{\mathrm{C},i}}{\sum_{k=1}^{K}z_{u,k}R_k^{\mathrm{C}}}\right)^{-1}\ge D_u,\ \ \forall u \tag{23b}$$

$$\sum_{u=1}^{U}\eta_{u,t}^{\mathrm{C},i}z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}\left(\boldsymbol{P}^i\right)\le R_k^{\mathrm{C}},\ \ \forall k,t \tag{23c}$$

$$\sum_{u=1}^{U}\left(\eta_{u,t}^{\mathrm{S},i}+\zeta_u\eta_{u,t}^{\mathrm{C},i}\right)z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}\left(\boldsymbol{P}^i\right)\le R_k^{\mathrm{S}},\ \ \forall k,t \tag{23d}$$

$$\eta_t^{\mathrm{L},i}+\eta_{u,t}^{\mathrm{S},i}+\eta_{u,t}^{\mathrm{C},i}=1,\ \ \forall u,t \tag{23e}$$

$$0\le\eta_t^{\mathrm{L},i},\ \eta_{u,t}^{\mathrm{S},i},\ \eta_{u,t}^{\mathrm{C},i}\le 1,\ \ \forall u,t \tag{23f}$$

$$\delta_{\mathrm{T}}\ge 3\epsilon_0 \tag{23g}$$

where i is the iteration index; Eq. (22) is the power allocation sub-problem and Eq. (23) denotes the data stream scheduling sub-problem. Next, we discuss the solutions to Eqs. (22) and (23).

5.4. Solution to Eq. (22)

It is not difficult to certify that Eq. (22) is nonconvex. According to Theorem 1 in Ref. [48], the problem in Eq. (22) can be solved iteratively using successive convex approximation techniques after

applying the Taylor expansion to Eqs. (22b), (22c), and (22d) [49]. Denoting the iteration index as j, the problem in Eq. (22) is reformulated as

$$\min_{\boldsymbol{P}^{i,j},\delta_{\mathrm{T}}} \ \delta_{\mathrm{T}} \tag{24a}$$

$$\text{s.t.} \ \frac{D_u}{\delta_{\mathrm{T}}} - \sum_{t=1}^{N_{\mathrm{T}}} \frac{\bar{R}_{u,t}^{\mathrm{UL}}\left(\boldsymbol{P}^{i,j}\middle|\boldsymbol{P}^{i,j-1}\right)}{F_{u,t}(\eta^{i-1})\bar{R}_{u,t}^{\mathrm{UL}}\left(\boldsymbol{P}^{i,j}\middle|\boldsymbol{P}^{i,j-1}\right)+1-\eta_t^{\mathrm{L},i-1}} \le 0, \quad \forall u \tag{24b}$$

$$\sum_{u=1}^{U} \eta_{u,t}^{\mathrm{C},i-1} z_{u,k} \tilde{R}_{u,t}^{\mathrm{UL}}\left[p_{u,t}^{i,j}, J_{u,k}\left(\bar{\boldsymbol{p}}_{u,t}^{i,j}\right)\middle|p_{u,t}^{i,j-1}, J_{u,k}\left(\bar{\boldsymbol{p}}_{u,t}^{i,j-1}\right)\right] \le R_k^{\mathrm{C}}, \quad \forall k,t \tag{24c}$$

$$\sum_{u=1}^{U}\left(\eta_{u,t}^{\mathrm{S},i-1}+\zeta_u\eta_{u,t}^{\mathrm{C},i-1}\right) z_{u,k} \tilde{R}_{u,t}^{\mathrm{UL}}\left[p_{u,t}^{i,j}, J_{u,k}\left(\bar{\boldsymbol{p}}_{u,t}^{i,j}\right)\middle|p_{u,t}^{i,j-1}, J_{u,k}\left(\bar{\boldsymbol{p}}_{u,t}^{i,j-1}\right)\right] \le R_k^{\mathrm{S}}, \quad \forall k,t \tag{24d}$$

$$0 \le p_{u,t}^{i,j} \le P_{\max}, \quad \forall u,t \tag{24e}$$

$$\delta_{\mathrm{T}} \ge 3\epsilon_0 \tag{24f}$$

where

$$F_{u,t}(\eta^{i-1}) = \frac{\eta_t^{\mathrm{L},i-1}}{R^{\mathrm{L}}} + \frac{\left(\eta_{u,t}^{\mathrm{S},i-1}+\zeta_u\eta_{u,t}^{\mathrm{C},i-1}\right)}{\sum_{k=1}^{K} z_{u,k}R_k^{\mathrm{S}}} + \frac{\eta_{u,t}^{\mathrm{C},i-1}}{\sum_{k=1}^{K} z_{u,k}R_k^{\mathrm{C}}} \tag{25}$$

$$\tilde{R}_{u,t}^{\mathrm{UL}}\left(\boldsymbol{P}^{i,j}\middle|\boldsymbol{P}^{i,j-1}\right) = (1-\gamma_{\mathrm{UL}})B\sum_{k=1}^{K} z_{u,k}\bar{R}_{u,k}^{\mathrm{UL}}\left[p_{u,t}^{i,j}, J_{u,k}\left(\bar{\boldsymbol{p}}_{u,t}^{i,j}\right)\middle|p_{u,t}^{i,j-1}, J_{u,k}\left(\bar{\boldsymbol{p}}_{u,t}^{i,j-1}\right)\right] \tag{26}$$

$$J_{u,k}(\bar{\boldsymbol{p}}_{u,t}) = \sum_{v=1,v\neq u}^{U} p_{v,t}\theta_{u,v,k}, \ \ \bar{\boldsymbol{p}}_{u,t} = \{p_{v,t}\}, \quad v\neq u \tag{27}$$

$$\begin{aligned}&\bar{R}_{u,k}^{\mathrm{UL}}\left[p_{u,t}, J_{u,k}(\bar{\boldsymbol{p}}_{u,t})\middle|p_{u,t}^0, J_{u,k}\left(\bar{\boldsymbol{p}}_{u,t}^0\right)\right]\\ &= \log_2\left[p_{u,t}\theta_{u,u,k}+J_{u,k}(\bar{\boldsymbol{p}}_{u,t})+\sigma^2\right] - \log_2\left[J_{u,k}\left(\bar{\boldsymbol{p}}_{u,t}^0\right)+\sigma^2\right]\\ &\quad - \frac{1}{\ln 2\left[J_{u,k}\left(\bar{\boldsymbol{p}}_{u,t}^0\right)+\sigma^2\right]}\left[J_{u,k}(\bar{\boldsymbol{p}}_{u,t}) - J_{u,k}\left(\bar{\boldsymbol{p}}_{u,t}^0\right)\right]\end{aligned} \tag{28}$$

$$\begin{aligned}&\tilde{R}_{u,k}^{\mathrm{UL}}\left[p_{u,t}, J_{u,k}(\bar{\boldsymbol{p}}_{u,t})\middle|p_{u,t}^0, J_{u,k}\left(\bar{\boldsymbol{p}}_{u,t}^0\right)\right]\\ &= \log_2\left[p_{u,t}^0\theta_{u,u,k}+J_{u,k}\left(\bar{\boldsymbol{p}}_{u,t}^0\right)+\sigma^2\right] - \frac{1}{\ln 2\left[p_{u,t}^0\theta_{u,u,k}+J_{u,k}\left(\bar{\boldsymbol{p}}_{u,t}^0\right)+\sigma^2\right]}\\ &\quad\times\left[p_{u,t}\theta_{u,u,k}-p_{u,t}^0\theta_{u,u,k}+J_{u,k}(\bar{\boldsymbol{p}}_{u,t}) - J_{u,k}\left(\bar{\boldsymbol{p}}_{u,t}^0\right)\right] - \log_2\left[J_{u,k}(\bar{\boldsymbol{p}}_{u,t})+\sigma^2\right]\end{aligned} \tag{29}$$

Subsequently, the solution to Eq. (24) is provided according to the following property:

Property 1: The problem in Eq. (24) is convex, and its optimal solution is a feasible solution to Eq. (22).

Proof: See Appendix A.

Property 1 shows that Eq. (24) can be solved using conventional convex optimization tools [46], which also indicates that the solution to Eq. (22) can be iteratively derived using the solution to Eq. (24). The detailed steps of this method are presented in Algorithm 1.

Algorithm 1. Power allocation algorithm for solving Eq. (22).

input: $K, M, U, R^{\mathrm{L}}, \left\{R_k^{\mathrm{S}}, R_k^{\mathrm{C}}\right\}\forall k, \boldsymbol{\theta}, z, \{D_u\}\forall u, N_{\mathrm{T}}, \epsilon_0, P_{\max}, \gamma_{\mathrm{UL}}, B, \eta^{i-1}, \boldsymbol{P}^{i-1}$

1: initialization: $\boldsymbol{P}^{i,0} = \boldsymbol{P}^{i-1}$, $\epsilon = 1\times10^{-2}$, $j = 1$

2: solve Eq. (24), denoting the optimal solution as $(\delta_{\mathrm{T}}^*, \boldsymbol{P}^*)$, setting $\boldsymbol{P}^{i,j} = \boldsymbol{P}^*, \delta_{\mathrm{T}}\left(\boldsymbol{P}^{i,j}, \eta^{i-1}\right) = \delta_{\mathrm{T}}^*$

3: **while** $\left|1 - \frac{\delta_{\mathrm{T}}(\boldsymbol{P}^{i,j-1}, \eta^{i-1})}{\delta_{\mathrm{T}}(\boldsymbol{P}^{i,j}, \eta^{i-1})}\right| > \epsilon$ **do**

4: $j = j + 1$

5: solve Eq. (24), denoting the optimal solution as $(\delta_{\mathrm{T}}^*, \boldsymbol{P}^*)$, setting $\boldsymbol{P}^{i,j} = \boldsymbol{P}^*, \delta_{\mathrm{T}}\left(\boldsymbol{P}^{i,j}, \eta^{i-1}\right) = \delta_{\mathrm{T}}^*$

output: $\boldsymbol{P}^{i,j}, \delta_{\mathrm{T}}\left(\boldsymbol{P}^{i,j}, \eta^{i-1}\right)$

5.5. Solution to Eq. (23)

The problem in Eq. (23) is nonconvex because Eq. (23b) is concave with respect to η^i. Furthermore, it can be solved using the Taylor expansion and successive convex approximation techniques [49]. Denoting the iteration index j, Eq. (23) is reformulated as follows:

$$\min_{\eta^{i,j},\delta_{\mathrm{T}}} \ \delta_{\mathrm{T}} \tag{30a}$$

$$\begin{aligned}&\text{s.t.} \ \frac{D_u}{\delta_{\mathrm{T}}}\\ &\quad - \sum_{t=1}^{N_{\mathrm{T}}}\left\{\frac{1}{F_{u,t}(\eta^{i,j-1}) + \frac{1-\eta_t^{\mathrm{L},i,j-1}}{\sum_{k=1}^{K} z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P}^i)}} - \frac{F_{u,t}(\eta^{i,j}) - F_{u,t}(\eta^{i,j-1}) - \frac{\eta_t^{\mathrm{L},i,j}-\eta_t^{\mathrm{L},i,j-1}}{\sum_{k=1}^{K} z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P}^i)}}{\left[F_{u,t}(\eta^{i,j-1}) + \frac{1-\eta_t^{\mathrm{L},i,j-1}}{\sum_{k=1}^{K} z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P}^i)}\right]^2}\right\}\\ &\quad \le 0, \ \ \forall u\end{aligned} \tag{30b}$$

$$\sum_{u=1}^{U} \eta_{u,t}^{\mathrm{C},i,j} z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}\left(\boldsymbol{P}^i\right) \le R_k^{\mathrm{C}}, \ \ \forall k,t \tag{30c}$$

$$\sum_{u=1}^{U}\left(\eta_{u,t}^{\mathrm{S},i,j}+\zeta_u\eta_{u,t}^{\mathrm{C},i,j}\right) z_{u,k}\hat{R}_{u,k,t}^{\mathrm{UL}}\left(\boldsymbol{P}^i\right) \le R_k^{\mathrm{S}}, \quad \forall k,t \tag{30d}$$

$$\eta_t^{\mathrm{L},i,j} + \eta_{u,t}^{\mathrm{S},i,j} + \eta_{u,t}^{\mathrm{C},i,j} = 1, \ \ \forall u,t \tag{30e}$$

$$0 \le \eta_t^{\mathrm{L},i,j}, \ \eta_{u,t}^{\mathrm{S},i,j}, \ \eta_{u,t}^{\mathrm{C},i,j} \le 1, \quad \forall u,t \tag{30f}$$

$$\delta_{\mathrm{T}} \ge 3\epsilon_0 \tag{30g}$$

We find that Eq. (30) is a convex optimization problem with respect to δ_{T} and $\eta^{i,j}$, and it can be solved using conventional convex optimization tools [50]. In addition, denoting the optimal solution to Eq. (30) as $(\delta_{\mathrm{T}}^*, \eta^*)$, it is not difficult to prove that $(\delta_{\mathrm{T}}^*, \eta^*)$ also belongs to the feasible region of Eq. (23). Consequently, the solution to Eq. (23) can be iteratively derived based on the solution to Eq. (30), as presented in Algorithm 2.

Algorithm 2. Data stream scheduling algorithm for solving Eq. (23).

input: $K, M, U, R^{\mathrm{L}}, \left\{R_k^{\mathrm{S}}, R_k^{\mathrm{C}}\right\}\forall k, \boldsymbol{\theta}, z, \{D_u\}\forall u, N_{\mathrm{T}}, \epsilon_0, P_{\max}, \gamma_{\mathrm{UL}}, B, \eta^{i-1}, \boldsymbol{P}^i$

1: initialization: $\eta^{i,0} = \eta^{i-1}, \epsilon = 1 \times 10^{-2}, j = 1$

2: solve Eq. (30), denoting the optimal solution as $(\delta_{\mathrm{T}}^*, \eta^*)$, setting $\eta^{i,j} = \eta^*, \delta_{\mathrm{T}}\left(\boldsymbol{P}^i, \eta^{i,j}\right) = \delta_{\mathrm{T}}^*$

3: **while** $\left|1 - \frac{\delta_{\mathrm{T}}\left(\boldsymbol{P}^i, \eta^{i,j}\right)}{\delta_{\mathrm{T}}\left(\boldsymbol{P}^i, \eta^{i,j-1}\right)}\right| > \epsilon$ **do**

4: $j = j + 1$

5: solve Eq. (30), denoting the optimal solution as $(\delta_{\mathrm{T}}^*, \eta^*)$, setting $\eta^{i,j} = \eta^*, \delta_{\mathrm{T}}\left(\boldsymbol{P}^i, \eta^{i,j}\right) = \delta_{\mathrm{T}}^*$

output: $\eta^{i,j}, \delta_{\mathrm{T}}\left(\boldsymbol{P}^i, \eta^{i,j}\right)$

After the problems in Eqs. (22) and (23) are solved, the solution to Eq. (17) can be iteratively derived by jointly using Algorithms 1 and 2 according to the block coordinate descent technique [8]. The detailed steps of the proposed joint resource orchestration scheme are summarized in Algorithm 3.

Algorithm 3. Joint resource orchestration algorithm for solving Eq. (17).

input: $K, M, U, R^{\mathrm{L}}, \left\{R_k^{\mathrm{S}}, R_k^{\mathrm{C}}\right\}\forall k, \boldsymbol{\theta}, z, \{D_u\}\forall u, N_{\mathrm{T}}, \epsilon_0, P_{\max}, \gamma_{\mathrm{UL}}, B$

1: initialization: $\boldsymbol{P} = 1 \times 10^{-3}\boldsymbol{U}_{U\times N_{\mathrm{T}}}$, where the elements in $\boldsymbol{U}_{U\times N_{\mathrm{T}}}$ are uniformly distributed random variables in the range of $[0, 1]$, $\epsilon = 1 \times 10^{-3}, i = 1$

2: solve Eq. (23) using Algorithm 2, denoting the optimal solution as $(\delta_{\mathrm{T}}^*, \eta^*)$, setting $\eta^0 = \eta^*, \delta_{\mathrm{T}}\left(\boldsymbol{P}^0, \eta^0\right) = \delta_{\mathrm{T}}^*$

3: **while** $\left|1 - \frac{\delta_{\mathrm{T}}\left(\boldsymbol{P}^i, \eta^{i,j}\right)}{\delta_{\mathrm{T}}\left(\boldsymbol{P}^i, \eta^{i,j-1}\right)}\right| > \epsilon$ **do**

4: solve Eq. (23) using Algorithm 2, denoting the optimal solution as $(\delta_{\mathrm{T}}^*, \eta^*)$ setting $\eta^i = \eta^*, \delta_{\mathrm{T}}\left(\boldsymbol{P}^i, \eta^i\right) = \delta_{\mathrm{T}}^*$

5: **while** $\left|1 - \frac{\delta_{\mathrm{T}}\left(\boldsymbol{P}^{i-1}, \eta^{i-1}\right)}{\delta_{\mathrm{T}}\left(\boldsymbol{P}^i, \eta^i\right)}\right| > \epsilon$ **do**

6: $i = i + 1$

7: solve Eq. (22) using Algorithm 1, denoting the optimal solution as $(\delta_{\mathrm{T}}^*, \boldsymbol{P}^*)$, setting $\boldsymbol{P}^i = \boldsymbol{P}^*, \delta_{\mathrm{T}}\left(\boldsymbol{P}^i, \eta^{i-1}\right) = \delta_{\mathrm{T}}^*$

8: solve Eq. (23) using Algorithm 2, denoting the optimal solution as $(\delta_{\mathrm{T}}^*, \eta^*)$, setting $\eta^i = \eta^*, \delta_{\mathrm{T}}\left(\boldsymbol{P}^i, \eta^i\right) = \delta_{\mathrm{T}}^*$

output: $\boldsymbol{P}^i, \eta^i, \delta_{\mathrm{T}}\left(\boldsymbol{P}^i, \eta^i\right)$

5.6. Solution to Eq. (18)

Comparing Eq. (18b) with Eq. (17b), we can state the following property:

Property 2: The optimal solution to Eq. (18) must satisfy $\eta_{u,t}^{\mathrm{C}} = 0, \forall u, t$.

Proof: It is observed that Eq. (18b) is uncorrelated with $\eta_{u,t}^{\mathrm{C}}$ such that the values of $\eta_{u,t}^{\mathrm{C}}$ do not influence the value of δ_{T} in Eq. (18b). Furthermore, if $\eta_{u,t}^{\mathrm{C}} = 0$ is substituted into Eq. (18), all constraints in Eqs. (18c)–(18h) can be satisfied. Consequently, $\eta_{u,t}^{\mathrm{C}} = 0$ always belongs to the feasible region of Eq. (18) for any u and t.

Using Property 2, Eq. (18) can be simplified to

$$\min_{\boldsymbol{P}, \eta, \delta_{\mathrm{T}}} \ \delta_{\mathrm{T}} \tag{31a}$$

$$\sum_{t=1}^{N_{\mathrm{T}}} \delta_{\mathrm{T}} \left(\frac{\eta_t^{\mathrm{L}}}{R^{\mathrm{L}}} + \frac{\eta_{u,t}^{\mathrm{S}}}{\sum_{k=1}^{K} z_{u,k} \hat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P})} + \frac{\eta_{u,t}^{\mathrm{S}}}{\sum_{k=1}^{K} z_{u,k} R_k^{\mathrm{S}}}\right)^{-1} \geq D_u, \quad \forall u \tag{31b}$$

$$\sum_{u=1}^{U} \eta_{u,t}^{\mathrm{S}} z_{u,k} \hat{R}_{u,k,t}^{\mathrm{UL}}(\boldsymbol{P}) \leq R_k^{\mathrm{S}}, \quad \forall k, t \tag{31c}$$

$$0 \leq p_{u,t} \leq P_{\max}, \quad \forall u, t \tag{31d}$$

$$\eta_t^{\mathrm{L}} + \eta_{u,t}^{\mathrm{S}} = 1, \quad \forall u, t \tag{31e}$$

$$0 \leq \eta_t^{\mathrm{L}}, \ \eta_{u,t}^{\mathrm{S}} \leq 1, \quad \forall u, t \tag{31f}$$

$$\delta_{\mathrm{T}} \geq 2\epsilon_0 \tag{31g}$$

which can be solved using Algorithms 1–3 with $\eta_{u,t}^{\mathrm{C}} = 0, \forall u, t$.

Remark 2: The most important difference between the solutions to Eqs. (17) and (18) is the possibility of using MEC for computing. Similar to the discussion in Remark 1, the communication and computing process may be accomplished with one segmentation when D_u is small, whereas a smaller overall communication and computing latency can be achieved if MEC is not used. The reason is that the overall latency could be lower than $3\epsilon_0$ if MEC is not used; however, the latency must be at least $3\epsilon_0$ if MEC is used, as expressed by Eqs. (17h) and (18h). This phenomenon can also be observed after the numerical results are derived.

Based on Algorithms 1–3, a process-oriented joint resource orchestration scheme to solve Eq. (13) is derived, as summarized in Algorithm 4. The minimum overall communication and computing latency can be obtained using Algorithm 4.

Algorithm 4. Proposed process-oriented joint resource orchestration algorithm.

input: $K, M, U, R^{\mathrm{L}}, \left\{R_k^{\mathrm{S}}, R_k^{\mathrm{C}}\right\}\forall k, \boldsymbol{\theta}, z, \{D_u\}\forall u, N_{\mathrm{T}}^{\mathrm{S}_1}, N_{\mathrm{T}}^{\mathrm{S}_2}, N_{\mathrm{T}}^{\mathrm{S}_3}, \epsilon_0, \epsilon_{\mathrm{a}}, P_{\max}, \gamma_{\mathrm{UL}}, B$

1: solve Eq. (16) with $N_{\mathrm{T}} = N_{\mathrm{T}}^{\mathrm{S}_1}$; then, $(\boldsymbol{P}^{\mathrm{S}_1}, \eta^{\mathrm{S}_1})$ is derived using Eq. (20) and $\delta_{\mathrm{T}}^{\mathrm{S}_1}$ is derived by Eq. (21)

2: solve Eq. (17) $N_{\mathrm{T}} = N_{\mathrm{T}}^{\mathrm{S}_2}$ using Algorithms 1–3. Denote the solution as $(\boldsymbol{P}^{\mathrm{S}_2}, \eta^{\mathrm{S}_2})$; then, $\delta_{\mathrm{T}}^{\mathrm{S}_2}$ is derived

3: solve Eq. (18) $N_{\mathrm{T}} = N_{\mathrm{T}}^{\mathrm{S}_3}$ using Algorithms 1–3. Denote the solution as $(\boldsymbol{P}^{\mathrm{S}_3}, \eta^{\mathrm{S}_3})$; then, $\delta_{\mathrm{T}}^{\mathrm{S}_3}$ is derived

4: calculate $T_{\min} = \min\left\{N_{\mathrm{T}}^{\mathrm{S}_1}\delta_{\mathrm{T}}^{\mathrm{S}_1}, N_{\mathrm{T}}^{\mathrm{S}_2}\delta_{\mathrm{T}}^{\mathrm{S}_2}, N_{\mathrm{T}}^{\mathrm{S}_3}\delta_{\mathrm{T}}^{\mathrm{S}_3}\right\} + \epsilon_{\mathrm{a}}$, where $(\boldsymbol{P}^*, \eta^*)$ is derived as the corresponding joint power allocation and data stream scheduling scheme

output: $T_{\min}, \boldsymbol{P}^*, \eta^*$

5.7. Convergence analysis

Here, we analyze the convergence of Algorithms 1–3. For the problem in Eq. (24) during the jth iteration step, we have

$$\delta_{\mathrm{T}}\left(\boldsymbol{P}^{i,j}, \eta^{i-1}\right) \leq \delta_{\mathrm{T}}\left(\boldsymbol{P}^{i,j-1}, \eta^{i-1}\right) \tag{32}$$

because both $\boldsymbol{P}^{i,j}$ and $\boldsymbol{P}^{i,j-1}$ are feasible solutions to Eq. (24) according to Property 1, and the minimum value of δ_{T} is achieved by $\boldsymbol{P}^{i,j}$. Thus, Algorithm 1 is guaranteed to converge according to Ref. [49], where we can derive $\boldsymbol{P}^{i,*}$ as the locally optimal solution that satisfies

$$\delta_{\mathrm{T}}\left(\boldsymbol{P}^{i,*},\eta^{i-1}\right)\leq\delta_{\mathrm{T}}\left(\boldsymbol{P}^{i,j-1},\eta^{i-1}\right),\ \boldsymbol{P}^{i,j}\rightarrow\boldsymbol{P}^{i,*},\ j\rightarrow\infty \tag{33}$$

Similarly, for the problem in Eq. (30) during the jth iteration step, we have

$$\delta_{\mathrm{T}}\left(\boldsymbol{P}^{i},\eta^{i,j}\right)\leq\delta_{\mathrm{T}}\left(\boldsymbol{P}^{i},\eta^{i,j-1}\right) \tag{34}$$

This indicates that Algorithm 2 is guaranteed to converge. Therefore, $\eta^{i,*}$ is the locally optimal solution that satisfies

$$\delta_{\mathrm{T}}\left(\boldsymbol{P}^{i},\eta^{i,*}\right)\leq\delta_{\mathrm{T}}\left(\boldsymbol{P}^{i},\eta^{i,j}\right),\ \eta^{i,j}\rightarrow\eta^{i,*},\ j\rightarrow\infty \tag{35}$$

According to Eqs. (32)–(35), we can conclude that

$$\delta_{\mathrm{T}}\left(\boldsymbol{P}^{i,*},\eta^{i,*}\right)\leq\delta_{\mathrm{T}}\left(\boldsymbol{P}^{i,*},\eta^{i-1,*}\right)\leq\delta_{\mathrm{T}}\left(\boldsymbol{P}^{i-1,*},\eta^{i-1,*}\right) \tag{36}$$

which shows that the objective function in Eq. (17a) continues to decrease when i increases. Owing to the constraints in Eqs. (17b)–(17h), the objective function must have a lower bound. Consequently, the convergence of Algorithm 3 is proven, and the locally optimal solution to Eq. (17) can be derived.

6. Simulation results and discussions

We used simulation results to evaluate the performance of the proposed algorithms. The parameters of the NTN are set as K = 7, M = 8, and U = 56, where U IoT devices are divided into K user groups, and each device is associated with the nearest UAV. The positions of the devices and UAVs are generated according to the discussions in Ref. [51], where the degree of user aggregation is set as $\beta=0.5$. The minimum distance between any two UAVs was set as $d_{\mathrm{UAV}}=30$ km, and the height of the UAV swarm was set as $h_k=3$ km, $\forall k$. We assume that the data sizes of all devices are the same, denoted as $D_u=D$, where we set D = 1 Gbit, $\forall u$, and the maximum transmission power of each device is set as $P_{\max}=2$ W [52]. As for the satellite system, the data rate of the device-satellite link is set as $R^{\mathrm{L}}=9.6$ kbit·s^{-1} [41], the maximum rate of the UAV–satellite link is set as $R_k^{\mathrm{S}}=2$ Mbit·s^{-1}, $\forall k$ [53], and the overall propagation time of the electromagnetic wave is set as $\epsilon_{\mathrm{a}}=240$ ms [54]. As for the UAV channel parameters, we set m = 4.02, $\Omega=25\times10^{-3}$, $\eta_{\mathrm{LOS}}=0.1$, $\eta_{\mathrm{NLOS}}=21$, f = 5.8 GHz, $c=3\times10^{8}$ m·s^{-1}, $\lambda=c/f$, $d_0=\lambda/2$, $a=5.0188$, and $b=0.3511$ [36,42]. As for the integrated communication and MEC system, we set $\zeta_u=0.01,\forall u$ [40], $\gamma_{\mathrm{UL}}=0.1$, B = 1 MHz, $N_{\mathrm{T}}^{s_1}=N_{\mathrm{T}}^{s_2}=N_{\mathrm{T}}^{s_3}=8$, $\sigma^2=-114$ dBm [45], and $\epsilon_0=500$ ms [54].

First, we evaluate the convergence performance of the proposed algorithms using numerical simulations, where we set $R_k^{\mathrm{C}}=10$ Mbit·s^{-1}, $\forall k$ [38]. Fig. 4 illustrates that Algorithm 1 converges after approximately five iterations, whereas Algorithm 2 only needs two iterations to converge, indicating that the data stream scheduling sub-problem is nearly a convex problem. In addition, because Algorithm 3 requires higher precision than the other algorithms, this algorithm requires more iterations to converge. Therefore, the proposed process-oriented scheme has a significant potential to be employed in practical systems.

Subsequently, we compared the proposed algorithm performance with other schemes. First, a simple scheme was considered, where we only used satellites for communication. Furthermore, the following three schemes were considered:

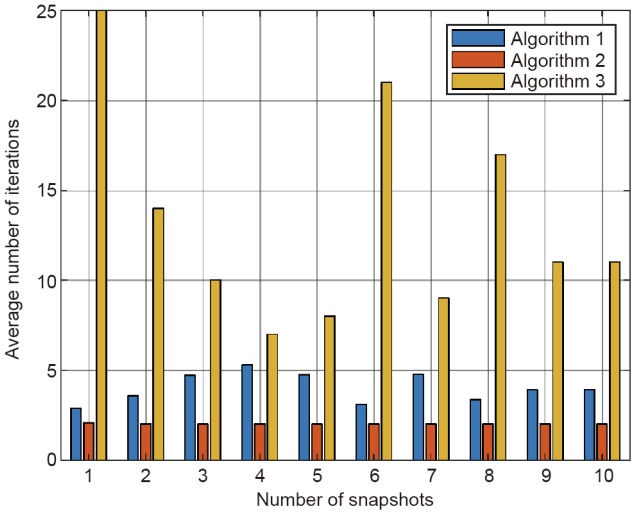

Fig. 4. Convergence performances of the proposed algorithms.

Scheme 1: We allocate the total bandwidth among multiple devices using the bandwidth allocation method proposed in Ref. [37], where the maximum transmission power is used and a greedy data stream scheduling strategy is applied, as presented in Ref. [10].

Scheme 2: The transmission power of each device is set to be equal [8], where power backoff is used to satisfy the constraints of the data stream and a greedy data stream scheduling strategy is applied, as presented in Ref. [10].

Scheme 3: A simplified version of the proposed algorithm is used, where we assume that $N_{\mathrm{T}}^{s_1}=N_{\mathrm{T}}^{s_2}=N_{\mathrm{T}}^{s_3}=1$ always holds.

In this simulation, we set $R_k^{\mathrm{C}}=6$ Mbit·s^{-1}, $\forall k$, and the performance of different algorithms were evaluated when the data sizes varied. According to Figs. 5 and 6, the proposed algorithm exhibits the highest performance in comparison to other algorithms; however, the performance gain varies with different data sizes. Fig. 5 illustrates that when D is less than 1 Mbit, the performance of the proposed algorithm is similar to that of Scheme 3 because the communication and computing process can be accomplished with one segmentation when the data size is small. In addition, a piecewise overall latency pattern can be observed when the proposed algorithm is used because the design of the proposed

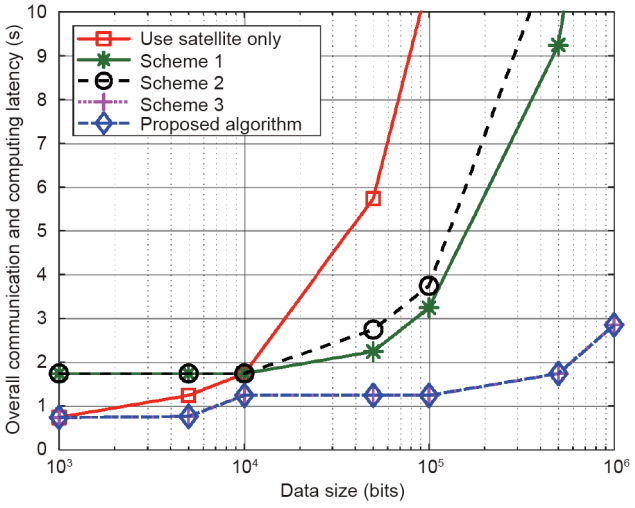

Fig. 5. Comparison between different algorithms when D is small.

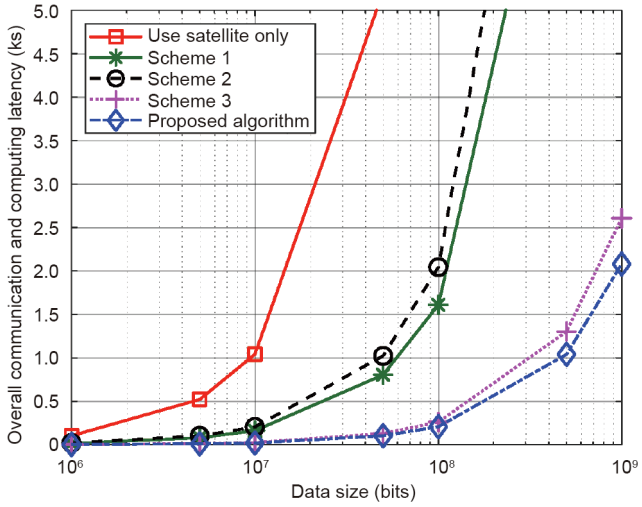

Fig. 6. Comparison between different algorithms when D is large.

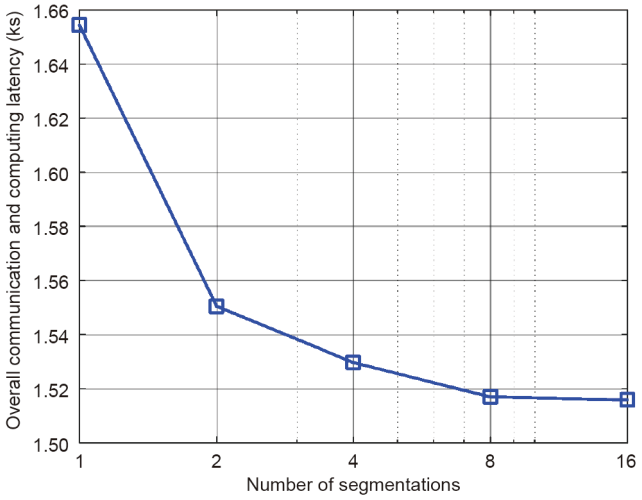

Fig. 7. Relationship between the minimum overall communication and computing latency and segmentation numbers.

scheme can adapt to varying data sizes. This phenomenon demonstrates the advantages of jointly designing the power allocation and data stream scheduling schemes. Furthermore, the curves in Fig. 6 demonstrate that a performance gain of approximately 30% is achieved using the proposed algorithm because N_{T} and δ_{T} can be appropriately designed using Algorithms 1–4. Some typical values of the overall communication and computing latency are presented in Table 2 to further clarify the characteristics of the proposed scheme, where the data sizes and update intervals vary. The minimum overall latencies with different data sizes are highlighted in bold that demonstrate the phenomenon described in Remarks 1 and 2. It can also prove that the proposed algorithm has the capability to adapt to different data sizes. Consequently, it is beneficial to use the proposed process-oriented scheme in a hierarchical NTN with MEC.

The curve in Fig. 7 is used to evaluate the minimum overall latency derived by the proposed algorithm when the process is designed with different numbers of segmentations, where we set $R_k^{\mathrm{C}} = 6$ Mbit·s^{-1}, $\forall k$ and $N_{\mathrm{T}}^{s_1} = N_{\mathrm{T}}^{s_2} = N_{\mathrm{T}}^{s_3} = N_{\mathrm{T}}$ as the segmentation number. We can observe that the latency performance can be improved by dividing the process into more segments. Moreover, the computational complexity increases with more segmentations; however, the performance gain decreases. Therefore, the value of the segmentation number should be appropriately selected to balance the computational complexity and performance of the proposed algorithm.

The curves in Fig. 8 are used to investigate the influence of the UAV payload on the latency performance of the proposed algorithm, where the average throughputs of MEC servers and data rates of UAV–satellite links vary. In this simulation, we set $R_k^{\mathrm{S}} = R^{\mathrm{S}}$ and $R_k^{\mathrm{C}} = R^{\mathrm{C}}$, $\forall k$. Fig. 8 shows that the overall latency is maintained constant with respect to R^{C} when R^{C} is smaller than R^{S}, because most data are scheduled to be sent back to the satellite through UAV–satellite links. In addition, the latency cannot be infinitely reduced by increasing R^{C} because limited data rates of device–UAV links may appear as a problem in this case. These phenomena imply that the payload deployment of UAVs should be appropriately designed prior to the communication and computing process as this could improve the efficiency of resource consumption regarding communications and MEC in a hierarchical NTN.

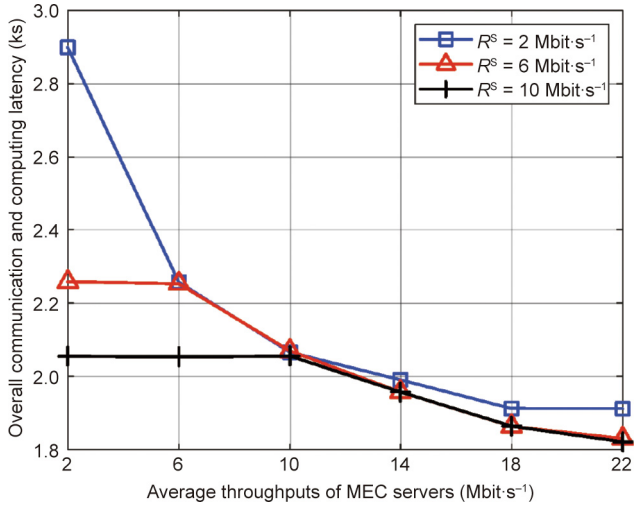

Fig. 8. Relationship between the overall communication and computing latency and average throughputs of MEC servers with varying data rates for UAV–satellite links.

Fig. 9 is used to discuss the influence of user distribution on the latency performance, where varying average throughputs of MEC servers are considered. In the simulation, we set $R_k^{\mathrm{C}} = R^{\mathrm{C}}$, $\forall k$. The curves show that decreased latencies can be achieved when β increases. This is because the proposed process-oriented scheme can slightly reduce the interference between neighboring devices, which further shows that the diversity in time is fully used under

Table 2
Overall communication and computing latencies with different data sizes.

Data size	Overall communication and computing latency derived by using the proposed scheme		
	Only use satellite for communication (s)	Use satellite and UAVs for communication without MEC (s)	Use satellite and UAVs for communication with MEC (s)
1 kbit	0.74	1.24	1.74
10 kbit	1.74	1.24	1.74
100 kbit	10.74	1.24	1.74
1 Mbit	104.74	5.24	3.24
10 Mbit	1.04×10^3	46.24	22.74
100 Mbit	1.04×10^4	451.24	213.24
1 Gbit	1.04×10^5	4.50×10^3	2.12×10^3

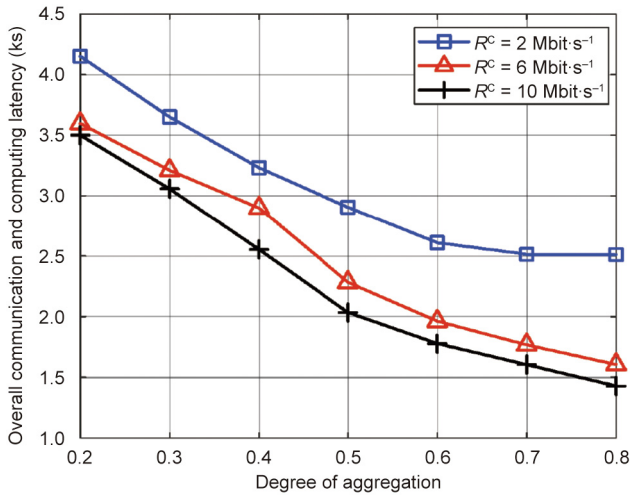

Fig. 9. Relationship between the overall communication and computing latency and degree of aggregation with different average throughputs of MEC servers.

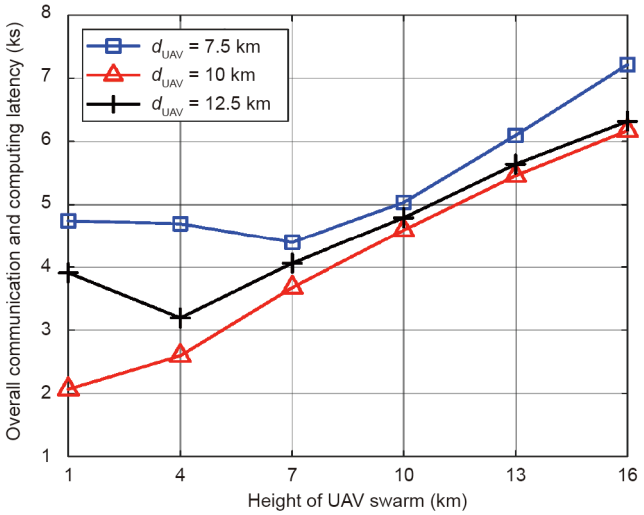

Fig. 11. Relationship between the overall communication and computing latency and height of the UAV swarm, where the distance between different UAVs varies.

the process-oriented framework. In particular, when $\beta \geq 0.7$ and $R^C = 2$ Mbit·s^{-1}, changing the degree of aggregation does not influence the latency performance because the limited throughput of MEC is the main problem for latency in this case. Moreover, the gaps between curves decrease when R^C continues to increase, as shown in Fig. 8.

Furthermore, we evaluated changes in the latency performance when different numbers of IoT devices were used, as shown in Fig. 10. In the simulation, we set $R_k^C = 6$ Mbit·s^{-1}, $\forall k$ and M = 16. It is observed that the overall latency is larger when more IoT devices are used because the resources of communication and MEC are limited to more devices. Moreover, the degree of aggregation has a greater influence on the latency performance with larger number of devices because the interferences between IoT devices have more chances to be increased. Therefore, the latency performance of the MEC-empowered NTN is sensitive to user distribution, particularly when the number of IoT devices is large in this network.

The curves in Fig. 11 are used to explore the relationship between the UAV positions and latency performance, where we set $R_k^C = 10$ Mbit·s^{-1}, $\forall k$. The overall latency does not always monotonically increase with respect to the UAV swarm height. This is because although the path loss is larger with higher UAVs, the interference between multiple devices could decrease, proving that higher altitudes of UAVs may have positive effects on the latency performance of algorithms. Therefore, the position and height of the UAV swarm should be appropriately designed in a hierarchical NTN to improve its latency performance.

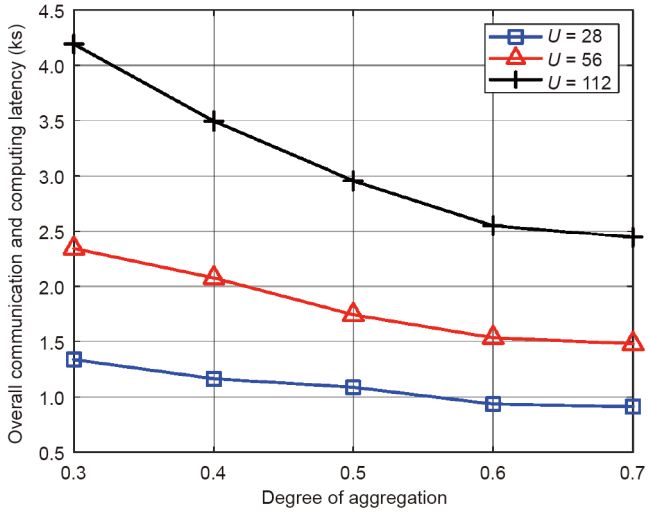

Fig. 10. Relationship between the overall communication and computing latency and degree of aggregation with different numbers of IoT devices.

7. Conclusions

In this study, the design of an MEC-empowered NTN for a wide-area time-sensitive IoT was investigated. To jointly design the communication and MEC systems for hierarchically integrated satellites and UAVs, a process-oriented framework was presented in a time-division manner. Under this framework, a latency minimization problem was formulated using the large-scale CSI. Subsequently, the problem could be transformed into a simplified form, and an approximation of the simplified problem was derived. The approximated problem was decomposed into sub-problems based on the properties of the overall communication and computing efficiency function. Additionally, an iterative algorithm was proposed to solve these sub-problems by jointly using block coordinate descent and successive convex approximation techniques. A process-oriented joint resource orchestration scheme was proposed for the MEC-empowered NTN using the solutions to the sub-problems. The simulation results demonstrated that the proposed process-oriented scheme exhibited a higher performance than that of the other comparison algorithms. In addition, simulations proved that the proposed process-oriented scheme could flexibly adapt to varying data sizes. Therefore, the payload deployments of UAVs should be appropriately predesigned to improve the efficiency of resource use in the MEC-empowered NTN. Finally, the results implied that it is advantageous to integrate NTN with MEC for wide-area time-sensitive IoT.

Acknowledgments

This work was supported in part by the National Key R&D Program of China (2018YFA0701601 and 2020YFA0711301), the National Natural Science Foundation of China (61771286, 61941104, and 61922049), and the Tsinghua University–China Mobile Communications Group Co., Ltd. Joint Institute.

Compliance with ethics guidelines

Chengxiao Liu, Wei Feng, Xiaoming Tao, and Ning Ge declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eng.2021.11.002.

References

[1] Saarnisaari H, Dixit S, Alouini MS, Chaoub A, Giordani M, Kliks A, et al. A 6G white paper on connectivity for remote areas. 2020. arXiv: 2004.14699.
[2] Huang C, Huang G, Liu W, Wang R, Xie M. A parallel joint optimized relay selection protocol for wake-up radio enabled WSNs. Phys Commun 2021;47:101320.
[3] FG-NET-2030. Network 2030: a blueprint of technology, applications and market drivers towards the year 2030 and beyond. Geneva: ITU; 2019.
[4] Wei T, Feng W, Chen Y, Wang CX, Ge N, Lu J. Hybrid satellite–terrestrial communication networks for the maritime Internet of Things: key technologies, opportunities, and challenges. IEEE Internet Things J 2021;8(11):8910–34.
[5] Li X, Feng W, Wang J, Chen Y, Ge N, Wang CX. Enabling 5G on the ocean: a hybrid satellite–UAV–terrestrial network solution. IEEE Wirel Commun 2020;27(6):116–21.
[6] Wang Y, Feng W, Wang J, Quek TQS. Hybrid satellite–UAV–terrestrial networks for 6G ubiquitous coverage: a maritime communications perspective. IEEE J Sel Areas Commun 2021;39(11):3475–90.
[7] Onireti O, Qadir J, Imran MA, Sathiaseelan A. Will 5G see its blind side? Evolving 5G for universal Internet access. In: Proceedings of the 2016 workshop on Global Access to the Internet for All; 2016 Aug; Florianopolis, Brazil. New York: Association for Computing Machinery; 2016. p. 1–6.
[8] Liu C, Feng W, Chen Y, Wang CX, Ge N. Cell-free satellite–UAV networks for 6G wide-area Internet of Things. IEEE J Sel Areas Commun 2021;39(4):1116–31.
[9] Zhao J, Gao F, Wu Q, Jin S, Wu Y, Jia W. Beam tracking for UAV mounted SatCom on-the-move with massive antenna array. IEEE J Sel Areas Commun 2018;36(2):363–75.
[10] Cheng X, Lyu F, Quan W, Zhou C, He H, Shi W, et al. Space/aerial-assisted computing offloading for IoT applications: a learning-based approach. IEEE J Sel Areas Commun 2019;37(5):1117–29.
[11] Raza U, Kulkarni P, Sooriyabandara M. Low power wide area networks: an overview. IEEE Commun Surv Tutor 2017;19(2):855–73.
[12] Centenaro M, Vangelista L, Zanella A, Zorzi M. Long-range communications in unlicensed bands: the rising stars in the IoT and smart city scenarios. IEEE Wirel Commun 2016;23(5):60–7.
[13] Lo Bello L, Steiner W. A perspective on IEEE time-sensitive networking for industrial communication and automation systems. Proc IEEE 2019;107(6):1094–120.
[14] Liang W, Zheng M, Zhang J, Shi H, Yu H, Yang Y, et al. WIA-FA and its applications to digital factory: a wireless network solution for factory automation. Proc IEEE 2019;107(6):1053–73.
[15] Luvisotto M, Pang Z, Dzung D. High-performance wireless networks for industrial control applications: new targets and feasibility. Proc IEEE 2019;107(6):1074–93.
[16] TR 38.824: Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC). 3GPP standard. France: 3GPP; 2019.
[17] TR 38.825: Study on NR industrial Internet of Things (IoT). 3GPP standard. France: 3GPP; 2019.
[18] TR 38.821: Solutions for NR to support non-terrestrial networks (NTN). 3GPP standard. France: 3GPP; 2020.
[19] Ghosh A, Maeder A, Baker M, Chandramouli D. 5G evolution: a view on 5G cellular technology beyond 3GPP Release 15. IEEE Access 2019;7:127639–51.
[20] De Sanctis M, Cianca E, Araniti G, Bisio I, Prasad R. Satellite communications supporting Internet of Remote Things. IEEE Internet Things J 2016;3(1):113–23.
[21] Cioni S, De Gaudenzi R, Del Rio Herrero O, Girault N. On the satellite role in the era of 5G massive machine type communications. IEEE Netw 2018;32(5):54–61.
[22] Zhen L, Qin H, Zhang Q, Chu Z, Lu G, Jiang J, et al. Optimal preamble design in spatial group-based random access for satellite-M2M communications. IEEE Wirel Commun Lett 2019;8(3):953–6.
[23] Zhang Q, Jiang M, Feng Z, Li W, Zhang W, Pan M. IoT enabled UAV: network architecture and routing algorithm. IEEE Internet Things J 2019;6(2):3727–42.
[24] Chakareski J. UAV-IoT for next generation virtual reality. IEEE Trans Image Process 2019;28(12):5977–90.
[25] Ranjha A, Kaddoum G. Quasi-optimization of uplink power for enabling green URLLC in mobile UAV-assisted IoT networks: a perturbation-based approach. IEEE Internet Things J 2021;8(3):1674–86.
[26] Huang M, Liu A, Xiong NN, Wu J. A UAV-assisted ubiquitous trust communication system in 5G and beyond networks. IEEE J Sel Areas Commun 2021;39(11):3444–58.
[27] Islambouli R, Sharafeddine S. Optimized 3D deployment of UAV-mounted cloudlets to support latency-sensitive services in IoT networks. IEEE Access 2019;7:172860–70.
[28] Zhang L, Ansari N. Latency-aware IoT service provisioning in UAV-aided mobile-edge computing networks. IEEE Internet Things J 2020;7(10):10573–80.
[29] Tan Z, Qu H, Zhao J, Zhou S, Wang W. UAV-aided edge/fog computing in smart IoT community for social augmented reality. IEEE Internet Things J 2020;7(6):4872–84.
[30] Tun YK, Park YM, Tran NH, Saad W, Pandey SR, Hong CS. Energy-efficient resource management in UAV-assisted mobile edge computing. IEEE Commun Lett 2021;25(1):249–53.
[31] Wang J, Liu K, Pan J. Online UAV-mounted edge server dispatching for mobile-to-mobile edge computing. IEEE Internet Things J 2020;7(2):1375–86.
[32] Guo J, Huang G, Li Q, Xiong NN, Zhang S, Wang T. STMTO: a smart and trust multi-UAV task offloading system. Inf Sci 2021;573:519–40.
[33] Zeng Y, Wu Q, Zhang R. Accessing from the sky: a tutorial on UAV communications for 5G and beyond. Proc IEEE 2019;107(12):2327–75.
[34] Liu J, Du X, Cui J, Pan M, Wei D. Task-oriented intelligent networking architecture for the space–air–ground–aqua integrated network. IEEE Internet Things J 2020;7(6):5345–58.
[35] Cao P, Liu Y, Yang C, Xie S, Xie K. MEC-driven UAV-enabled routine inspection scheme in wind farm under wind influence. IEEE Access 2019;7:179252–65.
[36] Chen Y, Feng W, Zheng G. Optimum placement of UAV as relays. IEEE Commun Lett 2018;22(2):248–51.
[37] Pan Y, Jiang H, Zhu H, Wang J. Latency minimization for task offloading in hierarchical fog-computing C-RAN networks. In: Proceedings of 2020 IEEE International Conference on Communications; 2020 Jun 7–11; Dublin, Ireland; 2020. p. 1–6.
[38] Wang JB, Yang H, Cheng M, Wang JY, Lin M, Wang J. Joint optimization of offloading and resources allocation in secure mobile edge computing systems. IEEE Trans Vehicular Technol 2020;69(8):8843–54.
[39] Wang P, Yao C, Zheng Z, Sun G, Song L. Joint task assignment, transmission, and computing resource allocation in multilayer mobile edge computing systems. IEEE Internet Things J 2019;6(2):2872–84.
[40] Sharma J, Choudhury T, Satapathy SC, Sabitha AS. Study on H.265/HEVC against VP9 and H.264: on space and time complexity for codecs. In: Proceedings of 2018 International Conference on Communication, Computing and Internet of Things; 2018 Feb 17–19; Chennai, India; 2018. p. 106–10.
[41] Dymond A, Billowes C, Lopianowski M. Trends and potential for the use of satellites for rural telecommunications in developing countries. In: Proceedings of International Conference on Rural Telecommunications; 1988 May 23–25; London, UK; 1988. p. 126–9.
[42] Khuwaja AA, Chen Y, Zhao N, Alouini MS, Dobbins P. A survey of channel modeling for UAV communications. IEEE Commun Surv Tutor 2018;20(4):2804–21.
[43] Du J, Xu W, Deng Y, Nallanathan A, Vandendorpe L. Energy-saving UAV-assisted multi-user communications with massive MIMO hybrid beamforming. IEEE Commun Lett 2020;24(5):1100–4.
[44] Ammari ML, Fortier P. Low complexity ZF and MMSE detectors for the uplink MU-MIMO systems with a time-varying number of active users. IEEE Trans Vehicular Technol 2017;66(7):6586–90.
[45] Cao P, Liu W, Thompson JS, Yang C, Jorswieck EA. Semidynamic green resource management in downlink heterogeneous networks by group sparse power control. IEEE J Sel Areas Commun 2016;34(5):1250–66.
[46] Rost P. Achievable net-rates in multi-user OFDMA with partial CSI and finite channel coherence. In: Proceedings of 2012 IEEE Vehicular Technology Conference (VTC Fall); 2012 Sep 3–6; Quebec, QC, Canada; 2012. p. 1–5.
[47] Khoshnevis B, Yu W, Lostanlen Y. Two-stage channel quantization for scheduling and beamforming in network MIMO systems: feedback design and scaling laws. IEEE J Sel Areas Comm 2013;31(10):2028–42.
[48] Liu C, Feng W, Tao X, Ge N. MEC-empowered non-terrestrial networks for 6G wide-area time-sensitive Internet of Things. 2021. arXiv: 1103.21907.
[49] Sun Y, Babu P, Palomar DP. Majorization–minimization algorithms in signal processing, communications, and machine learning. IEEE Trans Signal Process 2017;65(3):794–816.
[50] Boyd S, Vandenberghe L, editors. Convex optimization. Cambridge: Cambridge University Press; 2004.
[51] Mirahsan M, Schoenen R, Yanikomeroglu H. Hethetnets: heterogeneous traffic distribution in heterogeneous wireless cellular networks. IEEE J Sel Areas Comm 2015;33(10):2252–65.
[52] Zhao B, Ren G, Dong X, Zhang H. Spatial group based optimal uplink power control for random access in satellite networks. IEEE Trans Vehicular Technol 2020;69(7):7354–65.
[53] Peng F, Cardona ÁS, Shafiee K, Leung VCM. TCP performance evaluation over GEO and LEO satellite links between performance enhancement proxies. In: Proceedings of 2012 IEEE Vehicular Technology Conference (VTC Fall); 2012 Sept 3–6; Quebec, QC, Canada; 2012. p. 1–5.
[54] Luglio M, Roseti C, Zampognaro F. Transport layer optimization for cloud computing applications via satellite: TCP Noordwijk+. China Commun 2014;11(12):105–19.