remote sensing

MDPI

Article

Research on Resource Allocation Method of Space Information Networks Based on Deep Reinforcement Learning

Xiangli Meng [1,*], Lingda Wu [1] and Shaobo Yu [1]

Science and Technology on Complex Electronic System Simulation Laboratory, Space Engineering University, Beijing 101416, China; wld@nudt.edu.cn (L.W.); 18813182800@163.com (S.Y.)

* Correspondence: 10211193@bjtu.edu.cn; Tel.: +010-6636-4329

Received: 16 January 2019; Accepted: 18 February 2019; Published: 21 February 2019

Abstract: The space information networks (SIN) have a series of characteristics, such as strong heterogeneity, multiple types of resources, and difficulty in management. Aiming at the problem of resource allocation in SIN, this paper firstly establishes a hierarchical and domain-controlled SIN architecture based on software-defined networking (SDN). On this basis, the transmission, caching, and computing resources of the whole network are managed uniformly. The Asynchronous Advantage Actor-Critic (A3C) algorithm in deep reinforcement learning is introduced to model the process of resource allocation. The simulation results show that the proposed scheme can effectively improve the expected benefits of unit resources and improve the resource utilization efficiency of the SIN.

Keywords: space information networks; software-defined network; deep reinforcement learning; transmission resource; caching resource; computing resource

1. Introduction

At present, with the gradual deepening of space science exploration and the continuous development of space information technology, the construction of space information systems presents a state of explosive development. However, the construction of all kinds of spatial information systems is still separate, forming a situation of repeated construction and "chimney-like development". Various navigation, communication, remote-sensing, and other satellites occupy a large amount of orbital resources. When a single satellite system completes a given task, it will have more idle states, resulting in a waste of space resources [1]. The proposal of the space information network (SIN) provides a solution to the above problems. The SIN became a research hotspot in the global field [2].

The SIN is a network system that acquires, transmits, and processes spatial information in real time on various space platforms (such as synchronous satellites or mid-orbit satellites, stratospheric balloons, and manned or unmanned aerial vehicles) [3]. Compared with the ground network, the SIN plays an irreplaceable role in earth observation, emergency communication, air transportation, space TT&C, and the expansion of national strategic interests [4]. Compared with the traditional satellite network, the SIN has a series of characteristics such as complex structure, dynamic topology change, large cross-domain spatial scale, and so on. Therefore, we need to build an efficient SIN architecture to realize the effective allocation and management of multi-dimensional resources in the SIN, which is of great significance for the construction of the SIN [5].

Software-defined networking (SDN) is a new network architecture with data forwarding which is control-separated and software-programmable. SDN adopts a centralized control surface and a distributed forwarding surface. The control plane uses the developed control and forwarding

communication interface to centralize the control of network devices on the forwarding plane, while providing flexible programmable capabilities [6]. The core idea of SDN is applied to the SIN. The data plane and control plane of satellite are separated, such that the satellite mainly implements simple forwarding and hardware configuration functions, thus solving the disadvantages of complex design and high cost of satellite nodes. The resources of transmission, caching, and computing of the whole network are allocated by the controller, which can not only lighten the burden of satellite nodes, but also benefit the unified management of the whole network [7,8].

Deep reinforcement learning is a new research hotspot at present. It combines the perception ability of deep learning with the decision-making ability of reinforcement learning, and can realize direct control from original input to output. The Asynchronous Advantage Actor-Critic (A3C) algorithm is a deep reinforcement learning algorithm proposed by DeepMind in 2015 [9]. The A3C algorithm evaluates the output action. On the basis of using the Actor-Critic framework, the idea of asynchronous training is introduced, which effectively improves the training efficiency and reduces the training time [10]. The application of the A3C algorithm in the SIN can effectively solve the problem of dynamic allocation of the SIN resources, thereby improving the utilization efficiency of the SIN resources.

Current research on the SIN mainly focuses on architecture design and the routing algorithm. Relevant research institutes and scholars proposed to apply SDN technology to the construction of the SIN, but there is a lack of specific multi-dimensional resource allocation methods. The main features of this paper are as follows:

1. Based on the core idea of SDN, a hierarchical and domain-controlled SIN architecture is established. The overall network architecture and network control architecture are designed.
2. On the basis of the SDN-based SIN architecture, the transmission resources, caching resources, and computing resources in the SIN are unified. Among them, the transmission resource depends on the coverage time of low Earth orbit (LEO) satellite to users, the transmission state of geostationary orbit (GEO) data relay satellite, and the communication link state.
3. The dynamic allocation of multi-dimensional resources in the SIN is modeled mathematically. A SIN resource allocation method based on the A3C algorithm is proposed.
4. The expected benefits of unit resources under different conditions are simulated and analyzed. The simulation results show that the proposed scheme of unified management of transmission resources, caching resources, and computing resources has better expected benefits, and can effectively improve the efficiency of the SIN resources.

The rest of this article is arranged as follows: Section 2 analyzes the related research of the SIN and SDN-based SIN. Section 3 proposes an SDN-based SIN architecture and builds the system model. In the Section 4, the dynamic allocation of multi-dimensional resources in the SIN is modeled as a deep reinforcement learning process based on the algorithm of A3C. The Section 5 simulates and analyzes the scheme proposed in this paper. Section 6 summarizes and discusses the full text.

2. Related Work

2.1. Space Information Networks

The SIN is an important international scientific frontier and strategic commanding height in the world today. At present, the representative projects are the (Space Communications and Navigation (SCaN) of the National Aeronautics and Space Administration (NASA) [11], the Transformational Satellite Communication System (TSAT) of the United States (US) [12], and the Integrated Space Infrastructure for Global Communications (ISICOM) of Europe [13]. SCaN plans to divide the network system into a backbone network, access network, spacecraft intranet, and adjacent network, which can adequately meet the needs of future space communications in the United States. Based on this framework, there is no need to build a new communication infrastructure for emerging tasks,

which can effectively avoid duplication of construction. TSAT was proposed by the US Department of Defense in 2002. It consists of a space segment, terminal segment, and mission operation segment. The goal of TSAT is to adapt to the transformation of communication needs of the US military, break the communication bottleneck, and provide users with a secure and high-speed communication architecture. Although the TAST project eventually stopped, its overall architecture laid the foundation for the development of SIN technology. In September 2008, the European Union adopted a Security Council resolution on "Making Future European Space Policy", and the concept of ISICOM came into being. The ISICOM system consists of a space-based network and a terrestrial network. The goal is to establish an independent Internet Protocol (IP)-based communication network, which combines microwave and laser links to achieve broadcast services, emergency services, telemedicine, distance education services, and other services.

Through the investigation and analysis of the current research situation of the SIN, according to the way of networking, the current SIN architecture can be divided into three categories: satellite-earth network, space-based network, and space–net–Earth network. Its typical system and main characteristics are shown in Table 1 [14,15].

Table 1. Comparison of different space information network (SIN) architectures.

Architecture	Satellite–Earth Network	Space-based Network	Space–net–Earth Network
Typical system	**Civil:** Inmarsat, O3b, OneWeb, Intersat **Military:** WGS, MUOS	**Civil:** Iridium **Military:** AEHF	**Civil:** SCaN, ISICOM **Military:** TSAT
Ground	Global distributed ground station network	The system can operate independently of the ground station	The earth and the sky cooperate with each other; the ground network does not need the global distribution of stations
Inter-satellite networking	No	Yes	Yes
Equipment on satellite	Simple	Complex	Moderate
Difficulty of System Maintenance	Simple	Complex	Moderate
Technical complexity	Simple	Complex	Moderate
Construction cost	Low	High	Moderate

In conclusion, the space–net–Earth network architecture can make full use of the wide-area coverage ability and the abundant transmission and processing ability of the space-based network, and reduce the complexity and cost of the system technology; it is a more appropriate reference for the construction of the SIN.

Aiming at the resource scheduling problem of the SIN, the current research mainly focuses on the resource scheduling of the GEO data relay satellite. Adinolfi used a backtracking heuristic algorithm to solve the resource scheduling problem of the GEO data relay satellite for the European Space Station [16]. Rojanasoonthon studied the tracking and data relay satellite system (TDRSS) of the United States, and studied the scheduling problem with two visual time windows [17]. Gu analyzed the resource and task constraints in the scheduling process of the GEO data relay satellite, and established the scheduling model of the GEO data relay satellite [18]. The current research lacks research on the overall resource allocation method for different types of nodes in the SIN.

2.2. SDN-Based Space Information Networks

Based on the advantages of SDN technology, some scholars and research institutes proposed its application in the SIN. The related research is still in its infancy, mainly focusing on the research of architecture and routing algorithms. Researchers at the Centre National de la Recherche Scientifique (CNRS) and Université de Toulouse studied handover decision-making algorithms in satellite networks through SDN's programmability [19]. Joint researchers from the Polytechnic University of Catalonia and the Greek National Research Center are exploring the introduction of SDN technology into satellite networks. SDN technology is used to improve the satellite network infrastructure, so as to improve the joint service capability of ground and satellite networks and hybrid access service capability [20]. Researchers from Hughes Network Systems Inc. of the United States directly proposed

a software-defined satellite network (SDSN) architecture, and applied it to their SPACEWAY system (a new generation of broadband satellite communication system). By establishing modules and performance objects, the extended allocation of inter-satellite packet routing addresses and the resource management control in the controller were realized [21]. Reference [22] proposes a networking architecture for on-board switching systems based on SDN. SDN on-board switching system can effectively reduce the load of traditional on-board switching systems, optimize the utilization of satellite channel resources, and improve the quality of service support capability of satellite communication networks. References [23,24] analyzed the routing algorithm of SDN-based SIN.

SIN includes a large number of heterogeneous nodes such as satellites, caches, mobile edge computing (MEC) servers, and so on. The allocation of the SIN resources involves the allocation of multi-dimensional resources such as transmission, caching, and computing. It is necessary to make overall considerations to achieve the maximum effective use of the SIN resources.

3. System Model

In this section, we firstly establish an SDN-based SIN architecture. On this basis, this article takes the use of LEO communication satellites and GEO data relay satellites for information transmission as an example; the network model, satellite coverage and transmission model, communication link model, caching model, and computing model of the SIN are analyzed.

3.1. SDN-Based Space Information Network Architecture

3.1.1. Overall Networking Architecture

Based on the core idea of SDN, this paper establishes a hierarchical and domain-controlled SIN architecture, whose overall network architecture is shown in Figure 1.

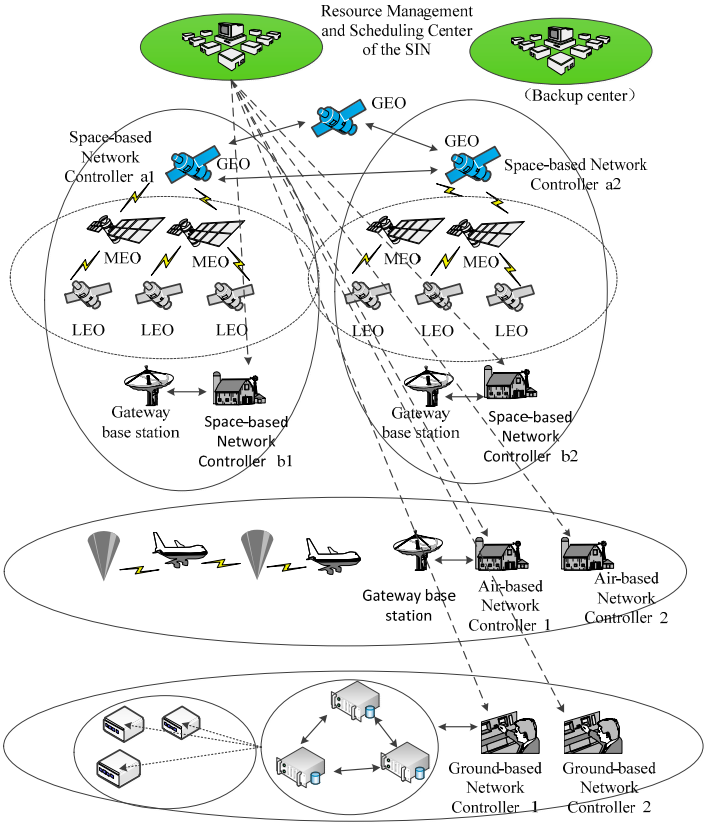

Figure 1. Overall networking architecture of the hierarchical and domain-controlled space information network (SIN) architecture.

From a hierarchical point of view, the SIN architecture is divided into three parts: space-based, air-based, and ground-based. The space-based network is mainly composed of satellites with different orbits, which are geostationary orbit satellites (GEO), medium orbit satellites (MEO), and low orbit satellites (LEO) from far to near. Air-based networks include stratospheric airships, balloons, manned or unmanned aerial vehicles, etc. Ground-based networks are mainly composed of gateway base stations, caches, and MEC servers, as well as space-based network controllers, space-based network controllers, ground-based network controllers, and an SIN resource management and scheduling center. Because the resource management and scheduling center of the SIN plays an important role, a backup center should be set up.

From the point of view of sub-domain, the ground-based network, space-based network, and air-based network are divided into several domains according to the region; each domain is controlled by a network controller. Among them, there are three kinds of network controllers, space-based network controllers, air-based network controllers, and ground-based network controllers, which control space-based networks, air-based networks, and ground-based networks, respectively. In order to make full use of the global coverage capability of GEO satellites and the high-speed computing capability of ground controllers, space-based network controllers are divided into space-based network controllers on the ground and space-based network controllers on the GEO satellite. Space-based network controllers on the ground are responsible for computing and storing large amounts of data and other complex functions. Space-based network controllers on the GEO satellite are responsible for collecting global views, completing simple routing storage, distributing flow tables, and other functions. Each network controller constitutes a single-domain controller, and multiple single-domain controllers are uniformly controlled by the SIN resource management and scheduling center [25].

3.1.2. Network Control Architecture

Based on the structure of SDN, the control architecture of the SIN is divided into three layers: application layer, control layer, and infrastructure layer [26]. The top layer is the application layer, which refers to a series of space tasks such as emergency communication and deep space exploration completed by the SIN. At the bottom is the infrastructure layer, which refers to satellites in different orbits, stratospheric vehicles, gateway base stations, and so on. In the middle is the control layer, which is composed of network controllers and the SIN resource management and scheduling center. The hierarchical and domain-based control structure of the SIN is shown in Figure 2.

In the control layer, the single-domain controller collects the topological information of each node in the domain. When the intra-domain traffic arrives, the single-domain controller calculates the intra-domain links, and controls the nodes by downloading the flow table, so as to realize path building and service processing. The SIN resource management and scheduling center is responsible for the control and allocation of the whole-network resources. It obtains the domain topology resources from the single-domain controller and establishes the whole-network topology. When cross-domain service arrives, it is responsible for cross-domain path calculation to realize cross-domain service transmission. In addition, due to the heterogeneity of different inter-domain networks, the SIN resource management and scheduling center is also responsible for the unification of heterogeneous device interfaces to achieve cross-domain interconnection of heterogeneous devices.

In the SIN management architecture based on SDN, north-direction agreement and south-direction agreement play an important role. North-direction agreement is a series of interfaces between application layer and control layer. There is no unified standard for its interface protocol. Therefore, the control layer provides many extensible application program interfaces (APIs) for different users in the application layer, and each API interface corresponds to a corresponding application; thus, the control architecture can implement a variety of application services. A typical south-direction agreement is OpenFlow [27], which is responsible for the interaction between the control layer and the underlying implementation switches to complete the forwarding of infrastructure

layer data. In the OpenFlow protocol, an OpenFlow switch can connect multiple network controllers; however, at the same time, only one controller has control over it, and other controllers have read-only function. In the SDN-based SIN management architecture, all switches in each single domain can only be managed by its single-domain controller.

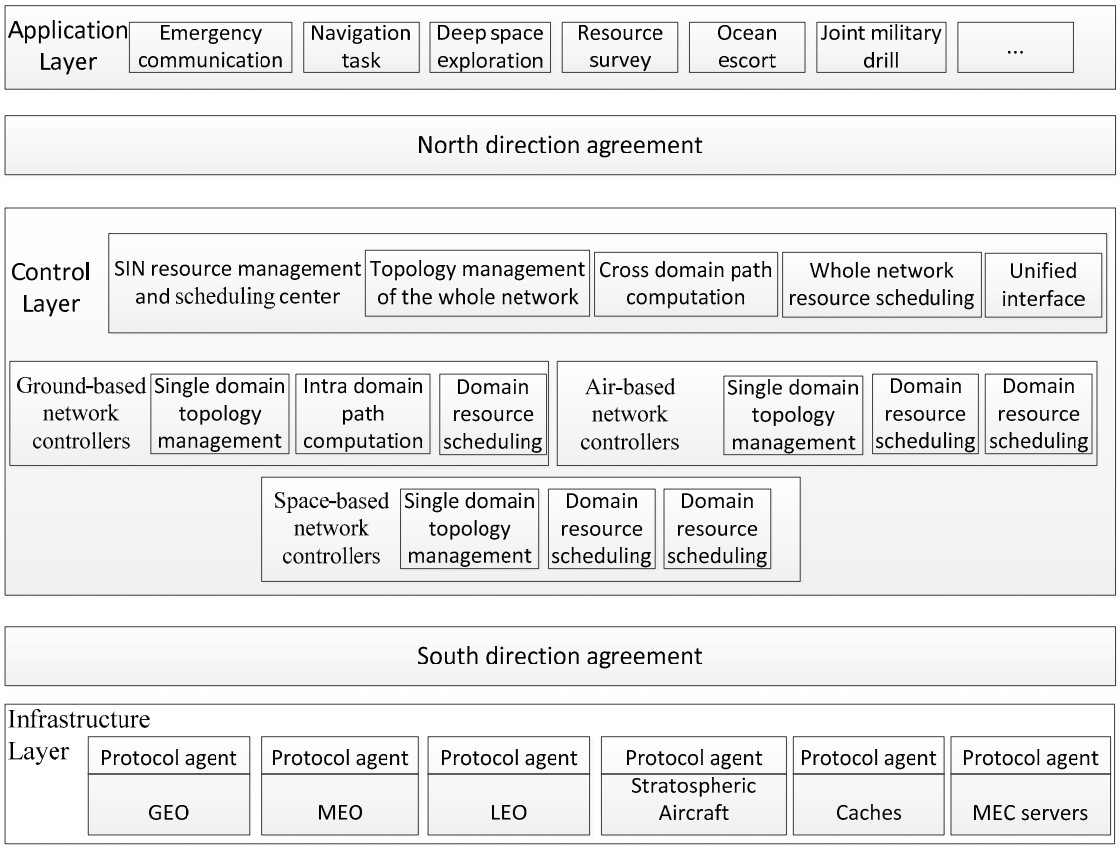

Figure 2. Network control architecture of the hierarchical and domain-controlled SIN architecture.

3.2. Network Model

The SIN resource management and scheduling center and the single-domain controllers realize the dispatching of various resources. This paper takes an LEO communication satellite and GEO data relay satellite as examples to analyze. Let *la*, *lga*, *ca*, *ma*, and *ua* represent the LEO communication satellite, GEO data relay satellite, cache device, MEC server, and user in the underlying physical resources, respectively. Let $la = \{1, \ldots, L\}$, $lga = \{1, \ldots, Lg\}$, $ca = \{1, \ldots, C\}$, $ma = \{1, \ldots, M\}$ and $ua = \{1, \ldots, U\}$, where L, Lg, C, M, and U represent the number of LEO satellites, GEO data relay satellites, caches, MEC servers, and users, respectively [28].

3.3. Satellite Coverage and Transmission Model

3.3.1. LEO Satellite Coverage Model

LEO satellite can only cover users in a certain time and space range to complete the transmission of information. The geometric relationship between LEO satellite and user is shown in Figure 3.

In Figure 3, O is the geocentric, R_e is the earth radius, h is the LEO satellite orbit altitude, and P represents the ground user; at t_0 time, the maximum elevation of the ground user is $\theta_{\max}$, and the LEO satellite position and the sub-satellite points are S' and M. At t time, the LEO satellite position and satellite sub-satellite points are S and N. Furthermore, $\gamma(t_0)$, $\gamma(t)$ and $\psi(t)$ represent the corresponding geocentric angles between P and M, P and N, and M and N, respectively. In addition, $\theta(t)$ represents

the elevation of the ground user at time t; $\theta(t)$ is the minimum elevation of the ground user, and the corresponding maximum geocentric angle at time t is $\gamma_{\max}$.

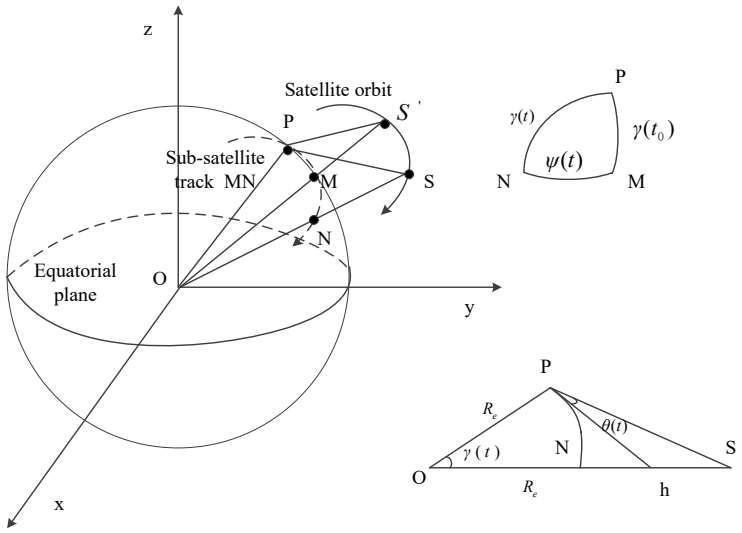

Figure 3. Geometric diagram of low Earth orbit (LEO) satellite and user.

According to the spherical triangle *PMN* and the triangle *OPS* shown in Figure 3, we can obtain the following:

$$\cos\gamma(t) = \cos\psi(t)\cos\gamma(t_0); \tag{1}$$

$$\gamma(t) = arc\ cos\left(\frac{R_e}{R_e + h}\cos\theta(t)\right) - \theta(t). \tag{2}$$

The effective coverage time t_c of LEO satellite to ground users is

$$t_c = \frac{2}{\omega}\psi(t) = \frac{2}{\omega}arc\ cos\left(\frac{\cos\gamma_{\max}}{\cos\gamma(t_0)}\right), \tag{3}$$

where $\omega = \omega_s - \omega_e i_0$ is the angular velocity of a satellite in the Earth-centered, Earth-fixed, (ECEF) coordinate system, ω_s is the angular velocity of a satellite in the Earth-centered inertial (ECI) coordinate system, ω_e is the angular velocity of the earth's rotation under ECI, and i_0 is the orbital inclination angle.

Ground users are randomly distributed. We assume that the distance from the ground user to the sub-satellite point obeys a uniform distribution. Therefore, when the LEO satellite covers ground users, $\gamma(t_0)$ satisfies the $U(0 \sim \gamma_{\max})$ uniform distribution. The probability density function $f_{\gamma(t_0)}(\gamma(t_0))$ of $\gamma(t_0)$ is

$$f_{\gamma(t_0)}(\gamma(t_0)) = \begin{cases} 1/\gamma_{\max}, 0 \le \gamma(t_0) < \gamma_{\max} \\ 0, \text{ Other} \end{cases}. \tag{4}$$

According to Equations (3) and (4), the cumulative distribution function of coverage time t_c is

$$\begin{aligned} &F_{T_c}(t_c) = P\left(\tfrac{2}{\omega}arc\ cos\left(\tfrac{\cos\gamma_{\max}}{\cos\gamma(t_0)}\right) \le t_c\right) = 1 - \tfrac{1}{\gamma_{\max}}arc\ cos\left(\cos\gamma_{\max}/\cos\tfrac{\omega t_c}{2}\right), \\ &0 < t_c \le T_m \end{aligned} \tag{5}$$

where T_m represents the maximum effective coverage time of satellite to ground users. When $\gamma(t_0) = 0$, according to Equation (3), we can get

$$T_m = \max(t_c) = 2\gamma_{\max}/\omega. \tag{6}$$

According to Equation (5), the probability density function of coverage time $f_{t_c}(t_c)$ is

$$f_{t_c}(t_c) = \begin{cases} \frac{\omega cosg_{max} tan(\omega t_c/2)}{2\gamma_{\max}\sqrt{\cos^2(\omega t_c/2) - \cos^2\gamma_{\max}}}, 0 < t_c \leq \frac{2\gamma_{\max}}{\omega} \\ 0 \text{ Other} \end{cases}. \quad (7)$$

According to Equation (7), the average coverage time $E(t_c)$ of LEO satellite to ground users is as follows [29]:

$$E(t_c) = \frac{2\cos\gamma_{\max}}{\omega\gamma_{\max}} \int_0^{\gamma_{\max}} \frac{x \tan x}{\sqrt{\cos^2 x - \cos^2\gamma_{\max}}} dx. \quad (8)$$

The elevation θ_u^l used between u and LEO satellite l is

$$\theta_u^l = arc\ tan\left(\frac{\cos\Theta - R_e/(R_e+h)}{\sin\Theta}\right), \quad (9)$$

where

$$cos\Theta = \cos(u_{lo} - l_{lo})\cos u_{la}\cos l_{la} + \sin u_{la}\sin l_{la}. \quad (10)$$

In Equation (10), u_{lo} and u_{la} represent the longitude and latitude of the user, respectively, while l_{lo} and l_{la} represent the longitude and latitude of LEO satellite, respectively.

When the LEO satellite is flying around the equator, l_{la} = 0, the longitude of the user and the satellite is the same, $u_{lo} = l_{lo}$, the elevation is the maximum, and Equation (10) can be simplified into

$$cos\Theta = \cos u_{la}. \quad (11)$$

Therefore, within the average coverage time $E(t_c)$, the maximum $\theta_{u\max}^l$ of θ_u^l is

$$\theta_{u\max}^l = arc\ tan(\frac{\cos u_{la} - R_e/(R_e+h)}{\sin u_{la}}). \quad (12)$$

To ensure that elevation increases monotonously, we set Ω as the elevation of LEO satellite from the horizon to the user. The relationship between Ω and θ_u^l is as follows:

$$\Omega = \begin{cases} \theta_u^l\ \theta_u^l \leq \theta_{u,\max}^l \\ 2 * \theta_{u,\max}^l - \theta_u^l\ \theta_u^l > \theta_{u,\max}^l \end{cases}. \quad (13)$$

The maximum value of Ω is $\Omega_{\max} = 2 * \theta_{u,\max}^l$. In this model, the smaller Ω is, the longer the LEO satellite coverage time will be. LEO satellites have more time to transmit, cache, and compute information with users. The larger Ω is, the shorter the coverage time of LEO satellite to users will be, and the less time it will take for the LEO satellite to transmit, cache, and compute information with users.

Because there are many LEO satellites in the SIN, we cannot determine which LEO satellite is connected to the user, nor can we determine the elevation of user u and satellite l at the next moment. Therefore, we set the elevation angle of user u and satellite l to the random variable Ω_u^l. The value range of Ω_u^l can be divided into Y' segments: $\Omega_0^* \leq \Omega_u^l \leq \Omega_1^*, y_0; \Omega_1^* \leq \Omega_u^l \leq \Omega_2^*, y_1; \ldots; \Omega_{Y'-1}^* \leq \Omega_u^l \leq \Omega_{u,\max}^l, y_{Y'-1}$. Each segment conforms to a Markov chain model and has Y' segments, that is, $y = \{y_0, y_1, \ldots, y_{Y'-1}\}$. The elevation of user u and LEO satellite l at time t is expressed as $w_u^l(t)$, where $t \in \{0, 1, 2, \ldots, T-1\}$. We have a total of T time slots, representing the total time from the user's application to the user's receiving and processing information. Based on a certain transition probability, $w_u^l(t)$ transfers from one state to another. The probability of transition from state $\overline{S11}$ to state $\overline{S12}$ is expressed as $\kappa_{\overline{S11S12}}(t)$. We can get a $Y' \times Y'$ dimensional elevation state transition probability matrix between user u and a LEO satellite l as follows:

$$\kappa_u^l(t) = \left[\kappa_{\overline{S11S12}}(t)\right]_{Y'\times Y'}, \quad (14)$$

where $\kappa_{\overline{S11S12}}(t) = \Pr\left(w_u^l(t+1) = \overline{S12} \middle| w_u^l(t) = \overline{S11}\right)$, $\overline{S11}, \overline{S12} \in y$.

3.3.2. GEO Data Relay Satellite Transmission Model

Due to the limited transmission capacity of the LEO communication satellite, it cannot meet the user's all-weather real-time transmission requirements. Therefore, the relay transmission mode of the GEO data relay satellite and LEO satellite will become an important part of the SIN [30].

We assume that the LEO satellite contains I tasks. Each task is arranged in descending order of importance. Task i represents the important task of the i-th item. The request rate of task i at time t is

$$\lambda_i(t) = \frac{\varpi}{\rho i^{\alpha}}. \tag{15}$$

The arrival process of task i obeys Poisson distribution with a ϖ parameter. The content of task request satisfies Zipf-like distribution. The probability of task i is $1/\rho i^{\alpha}$, where $\rho = \sum_{i=1}^{I} 1/i^{\alpha}$, α is the Zipf slope, and $0 < \alpha \leq 1$ [31].

We are not sure if task i requires the transmission of a GEO data relay satellite. Therefore, we assume that task i is transmitted by the GEO relay satellite as a random variable $\wp_i$. If task i does not require relay satellite transmission, then $\wp_i = 0$; otherwise, $\wp_i = 1$, constituting a Markov chain model $\wp_i = \{0, 1\}$ with two states. The transmission state of time t can be expressed as $\wp_i(t)$, $t \in \{0, 1, 2, \ldots, T-1\}$. According to a certain transition probability, the transmission state $\wp_i(t)$ is transferred from one state to another state. Let $J_{\overline{S21S22}}(t)$ denote the probability of transition from state $\overline{S21}$ to state $\overline{S22}$; then, the transition probability matrix $\diamond_i(t)$ is obtained as follows:

$$\diamond_i(t) = \left[J_{\overline{S21S22}}(t)\right]_{2\times 2}, \tag{16}$$

where $J_{\overline{S21S22}}(t) = \Pr\left(\wp_i(t+1) = \overline{S22} | \wp_i(t) = \overline{S21}\right)$, $\overline{S21}, \overline{S22} \in \wp_i$ [32].

3.4. Communication Link Model

According to Reference [33], the main models of satellite communication channel are the C. Loo model, Corazza model, and Lutz model. The C. Loo model is mainly suitable for rural environments. The received signals are mainly composed of direct shadowing signal components and multi-path signal components which are not shadowed. The Corazza model is applicable to all environments (roads, villages, cities, etc.). The signals received by users are affected by shadows. The Lutz model divides the channel environment between satellite and user into good and bad states. In the good state, there is no shadowing effect. In the bad state, there is no direct signal component. The above three models are represented as model X, model Y, and model Z, respectively. Three main propagation models of satellite communication links are shown in Figure 4.

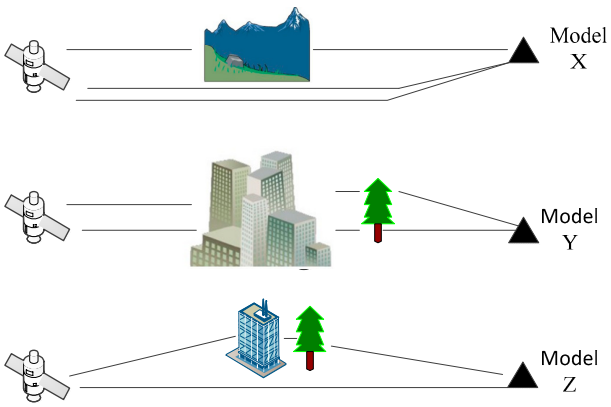

Figure 4. Satellite channel model diagram.

We assume that the probability of satellite link transmission models X, Y, and Z are p_X, p_Y and p_Z, respectively. From this, we get a three-element model space $S = \{S_X, S_Y, S_Z\}$. The state transition probability matrix Λ between the three models is

$$\begin{aligned}\Lambda &= \begin{pmatrix} P_{XX}\ P_{XY}\ P_{XZ} \\ P_{YX}\ P_{YY}\ P_{YZ} \\ P_{ZX}\ P_{ZY}\ P_{ZZ} \end{pmatrix} \\ &= \begin{pmatrix} 1-\frac{\Delta t}{<\Gamma_X>}\frac{\Delta t}{2p_X}\left(\frac{p_X}{<\Gamma_X>}+\frac{p_Y}{<\Gamma_Y>}-\frac{p_Z}{<\Gamma_Z>}\right) & \frac{\Delta t}{2p_X}\left(\frac{p_X}{<\Gamma_X>}+\frac{p_Z}{<\Gamma_Z>}-\frac{p_Y}{<\Gamma_Y>}\right) \\ \frac{\Delta t}{2p_Y}\left(\frac{p_X}{<\Gamma_X>}+\frac{p_Y}{<\Gamma_Y>}-\frac{p_Z}{<\Gamma_Z>}\right) & 1-\frac{\Delta t}{<\Gamma_X>}\frac{\Delta t}{2p_Y}\left(\frac{p_Y}{<\Gamma_Y>}+\frac{p_Z}{<\Gamma_Z>}-\frac{p_X}{<\Gamma_X>}\right) \\ \frac{\Delta t}{2p_Z}\left(\frac{p_X}{<\Gamma_X>}+\frac{p_Z}{<\Gamma_Z>}-\frac{p_Y}{<\Gamma_Y>}\right) & \frac{\Delta t}{2p_Z}\left(\frac{p_Y}{<\Gamma_Y>}+\frac{p_Z}{<\Gamma_Z>}-\frac{p_X}{<\Gamma_X>}\right) \end{pmatrix} \\ &\quad \begin{matrix} \\ \\ \end{matrix} \begin{matrix} \\ \\ 1-\frac{\Delta t}{<\Gamma_X>} \end{matrix}\end{aligned} \tag{17}$$

where Δt is the smallest unit of time for state transition between two transmission models, and $< \Gamma_X >$, $< \Gamma_Y >$ and $< \Gamma_Z >$ represent the average time of model states X, Y and Z, respectively.

We assume that the transmission link between satellite and user is time-varying and can be modeled as a finite-state Markov chain model. In this model, the quality of the channel is expressed as the signal-to-noise ratio (SNR) of the signal received by the user. We assume that the SNR of the signal received by user u from LEO satellite l is the random variable h_u^l. The value range of h_u^l can be divided into L' segments: $h_0^* \le h_u^l \le h_1^*, H_0; h_1^* \le h_u^l \le h_2^*, H_1; \ldots; h_u^l \ge h_{L'-1}^*, H_{L'-1}$. Each segment conforms to a Markov chain model and has L' segments, that is, $H = \{H_0, H_1, \ldots, H_{L'-1}\}$. At time t, the SNR of the signal received by user u from LEO satellite l is $h_u^l(t)$, where $t \in \{0, 1, 2, \ldots, T-1\}$. According to a certain transition probability, the SNR $h_u^l(t)$ is transferred from one state to another state. Let $\gamma_{\overline{S31}\overline{S32}}(t)$ denote the probability of transition from state $\overline{S31}$ to state $\overline{S32}$. The state transition probability matrix of transmission channel between user u and LEO satellite l can be expressed as an $L' \times L'$ dimensional matrix $l_u^l(t)$.

$$l_u^l(t) = \left[\gamma_{\overline{S31}\overline{S32}}(t)\right]_{L' \times L'}, \tag{18}$$

where $l_{\overline{S31}\overline{S32}}(t) = Pr\left(h_u^l(t+1) = \overline{S32} \middle| h_u^l(t) = \overline{S31}\right)$, $\overline{S31}, \overline{S32} \in H$.

We assume that the available spectrum bandwidth of the LEO satellite l is B^l Hz, where B_u^l Hz is allocated to user u. The available return capacity of satellite l is Z^l bps. User u's spectrum utilization at time t is $v_u^l(t)$. Then, the communication rate between user u and LEO satellite l is

$$ComR_u^l(t) = a_u^l(t) B_u^l(t) v_u^l(t), \forall u \in ua, \tag{19}$$

and $\sum_{u \in ua} ComR_u^l(t) \le Z^l, \forall l \in la$, where $a_u^l(t)$ indicates whether user u is connected to LEO satellite l. $a_u^l(t) = 1$ indicates that user u is connected to LEO satellite l; otherwise, $a_u^l(t) = 0$.

3.5. Caching Model

Based on the analysis of Section 3.3.2, users in the SIN have I tasks. Each task is arranged in descending order of importance. Task i represents the important task of the i-th item. The request rate of task i at time t is shown in Equation 15. The arrival process of task i obeys Poisson distribution with a ϖ parameter. The content of the task request satisfies a Zipf-like distribution. The probability of task i is $1/\rho i^{\alpha}$, where $\rho = \sum_{i=1}^{I} 1/i^{\alpha}$, α is a Zipf slope, and $0 < \alpha \le 1$ [34].

We cannot determine whether task i is cached first. Therefore, we assume that task i is cached as a random variable ς_i. If task i is not cached, then $\varsigma_i = 0$; otherwise, $\varsigma_i = 1$, constituting a Markov chain model $\varsigma_i = \{0, 1\}$ with two states. The cache state of time t can be expressed as $\varsigma_i(t)$,$t \in \{0, 1, 2, \ldots, T-1\}$. According to a certain transition probability, the cache state $\varsigma_i(t)$ is

transferred from one state to another state. Let $J_{\overline{S41S42}}(t)$ denote the probability of transition from state $\overline{S41}$ to state $\overline{S42}$; then, the transition probability matrix $\Phi_i(t)$ is obtained as follows:

$$\Phi_i(t) = \left[J_{\overline{S41S42}}(t)\right]_{2\times 2}, \tag{20}$$

where $J_{\overline{S41S42}}(t) = \Pr\left(\varsigma_i(t+1) = \overline{S42} | \varsigma_i(t) = \overline{S41}\right)$, $\overline{S41}, \overline{S42} \in \varsigma_i$ [35].

3.6. Computing Model

Let user u have computing task $T_u = \{o_u, n_u\}$, where o_u represents the size of the task content, and n_u represents the number of cycles that the central processing unit (CPU) needs to run to complete the task. Because there are multiple users and MEC servers, it is impossible to know how much computing power is allocated to user u. Therefore, a random variable Ξ_u^m is established to represent the computing power of assigning MEC server m to user u. Ξ_u^m is divided into M' discrete intervals, $\Pi = \{\Pi_0, \Pi_1, \ldots, \Pi_{M'-1}\}$. The computing state of time t can be expressed as $\Xi_u^m(t)$, $t \in \{0, 1, 2, \ldots, T-1\}$. According to a certain transition probability, the computing state $\Xi_u^m(t)$ is transferred from one state to another state. Let $\varepsilon_{\overline{S51S52}}(t)$ denote the probability of transition from state $\overline{S51}$ to state $\overline{S52}$. The state transition probability matrix $E_u^m(t)$ of $M' \times M'$ dimension can be expressed as

$$E_u^m(t) = \left[\varepsilon_{\overline{S51S52}}(t)\right]_{M'\times M'}, \tag{21}$$

where $\varepsilon_{\overline{S51S52}}(t) = \left(P_r \Xi_u^m(t+1) = \overline{S52} | \Xi_u^m(t+1) = \overline{S51}\right)$, $\overline{S51}, \overline{S52} \in \Pi$.

The execution time of task T_u on MEC server m is

$$t_u^m = \frac{n_u}{\Xi_u^m(t)}. \tag{22}$$

Thus, the computing rate is

$$CompR_u^m(t) = a_u^m(t)\frac{o_u}{t_u^m} = a_u^m(t)\frac{\Xi_u^m(t) o_u}{n_u}, \tag{23}$$

and $\sum_{u\in ua} a_u^m(t) o_u \le O_m$, where $a_u^m(t)$ indicates whether the user uses the MEC server m. $a_u^m(t) = 1$ means that the user uses MEC server m; otherwise, $a_u^m(t) = 0$. O_m represents the maximum value that can be calculated on server m [36].

4. Problem Equation

Based on the satellite coverage and transmission model, communication link model, caching model, and computing model established in Section 3, this section models the allocation of multi-dimensional resources in the SIN as a deep reinforcement learning process. Next, the state set, action set, reward function, and A3C algorithm flow in the process of deep reinforcement learning are analyzed.

4.1. State Set

The state set of the SIN includes the elevation state between user and satellite, transmission state of GEO data relay satellite, communication link state, caching state, and computing state. Therefore, the state set $S(t)$ of time t can be expressed as

$$S(t) = \begin{bmatrix} w_u^1(t)\ \mathrm{w}_u^2(t)\ \ldots\ \mathrm{w}_u^L(t) \\ \Re_l^1(t)\ \Re_l^2(t)\ \ldots\ \Re_l^{Lg}(t) \\ h_u^1(t)\ h_u^2(t)\ \ldots\ h_u^L(t) \\ \Gamma_u^1(t)\ \Gamma_u^2(t)\ \ldots\ \Gamma_u^C(t) \\ \Xi_u^1(t)\ \Xi_u^2(t)\ \ldots\ \Xi_u^M(t) \end{bmatrix}, \tag{24}$$

where $\Gamma_u^c(t) = [\varsigma_1(t), \varsigma_2(t), \ldots, \varsigma_i(t), \varsigma_I(t)]$, $\varsigma_i(t) \in [0,1]$, $\Re_l^{Lg}(t) = [\wp_1(t), \wp_2(t), \ldots, \wp_i(t), \wp_I(t)]$, and $\wp_i(t) \in [0,1]$.

4.2. Action Set

In the dynamic change of the SIN, we use a deep reinforcement learning algorithm to decide which LEO satellite is connected to user u, whether the tasks of user u need GEO data relay satellite for transmission, whether the tasks of user u are cached, and which MEC server is used to compute the tasks of user u. Therefore, the set of actions at time t is

$$a_u(t) = \{ComA_u(t), ComA_l(t), CaA_u(t), CompA_u(t)\}, \tag{25}$$

where the following apply:

(1) $ComA_u(t) = [ComA_u^1(t), ComA_u^2(t), \ldots, ComA_u^l(t), ComA_u^L(t)]$,$ComA_u^l(t) \in \{0,1\}$. When $ComA_u^l(t) = 0$, it means that user u is not connected to LEO satellite l at time t; otherwise, $ComA_u^l(t) = 1$. In this paper, at any time, it is assumed that only one LEO satellite is connected to the user u; thus, $\sum_{l \in la} ComA_u^l(t) = 1, \forall u \in ua$.

(2) $ComA_l(t) = [ComA_l^1(t), ComA_l^2(t), \ldots, ComA_l^{lg}(t), ComA_l^{Lg}(t)]$,$ComA_l^{lg}(t) \in \{0,1\}$. When $ComA_l^{lg}(t) = 0$, it means that the task is not transmitted by GEO data relay satellite lg; otherwise, $ComA_l^{lg}(t) = 1$. In this paper, at any time, it is assumed that only one GEO data relay satellite is connected to the LEO satellite; thus, $\sum_{lg Ilg a} ComA_l^{lg}(t) = 1, \forall l \in la$.

(3) $CaA_u(t) = [CaA_u^1(t), CaA_u^2(t), \ldots, CaA_u^c(t), CaA_u^C(t)]$,$CaA_u^c(t) \in \{0,1\}$. When $CaA_u^c(t) = 0$, it means that the task is not cached by cache c; otherwise, $CaA_u^c(t) = 1$. In this paper, at any time, suppose there is only one cache to cache a specified task; thus, $\sum_{c \in ca} CaA_u^c(t) = 1, \forall u \in ua$.

(4) $CompA_u(t) = [CompA_u^1(t), CompA_u^2(t), \ldots, CompA_u^m(t), CompA_u^M(t)]$,$CompA_u^m(t) \in \{0,1\}$. When $CompA_u^m(t) = 0$, it means that the task was not handed over to MEC server m for computing; otherwise, $CompA_u^m(t) = 1$. In this paper, at any time, it is supposed that there is only one MEC server to compute a specified task; thus, $\sum_{m \in ma} CompA_u^m(t) = 1, \forall u \in ua$.

4.3. Reward Function

According to Reference [37], SDN managers of the SIN need to pay for LEO satellite l, GEO data relay satellite lg, cache c, and MEC server m. It is assumed to pay δ_l to the LEO satellite every Hz, δ_{lg} to the GEO data relay satellite per Hz, ς_c to the cache per unit storage space, and η_m to the MEC server per joule.

In addition, the SIN managers need to charge users for information transmission, caching, and computing. Suppose τ_u is charged per bit of transmission information, κ_u is charged per bit of cache information, and ϕ_u is charged per bit of calculation information. The reward function is

$$\begin{aligned}
R_u(t) &= \sum_{l \in la} R_{u,l}^{comm}(t) + \sum_{lg \in lg a} R_{l,lg}^{comm}(t) + \sum_{c \in ca} R_{u,c}^{coche}(t) + \sum_{m \in ma} R_{u,m}^{comp}(t) \\
&= \sum_{l \in la} w_u^l(t) ComA_u^l(t) (\tau_u ComR_u^l(t)/\delta_l B_u^l(t)) + \sum_{\lg \in \lg a} w_u^l(t) ComA_l^{\lg}(t) (\tau_u ComR_l^{lg}(t)/\delta_{lg} B_l^{lg}(t)) \\
&+ \sum_{c \in ca} w_u^l(t) CaA_u^c(t) (\kappa_u CaR_u^c(t)/\varsigma_c o_u) + \sum_{m \in ma} w_u^l(t) CompA_u^m(t) (\phi_u CompR_u^m(t)/\eta_m n_u e_m) \\
&= \sum_{l \in la} w_u^l(t) ComA_u^l(t) (\tau_u B_u^l(t) v_u^l(t)/\delta_l B_u^l(t)) + \sum_{\lg \in \lg a} w_u^l(t) ComA_l^{\lg}(t) (\tau_u B_l^{lg}(t) v_l^{lg}(t)/\delta_{lg} B_l^{lg}(t)) \\
&+ \sum_{c \in ca} w_u^l(t) CaA_u^c(t) (\kappa_u B_u^l(t) v_u^l(t) \varsigma_u^c(t)/\varsigma_c o_u) + \sum_{m \in ma} w_u^l(t) CompA_u^m(t) (\phi_u \tfrac{\Xi_u^m(t) o_u}{n_u}/\eta_m n_u e_m)
\end{aligned} \tag{26}$$

where e_m represents the energy consumed by the CPU to rotate a circle. We define the reward function $R_u(t)$ as the expected benefit of the unit resource at time t, that is, the ratio of the fee charged to the user and the fee paid to obtain the resource. The higher the value of $R_u(t)$ is, the higher the utilization rate of resources will be.

4.4. A3C Algorithm

In this paper, we need to consider the coverage of the LEO satellite, transmission status of the GEO data relay satellite, communication link status, cache status, and computing power of the MEC server. Moreover, the SIN is a dynamic network system which is constantly changing. Therefore, this paper adopts the A3C algorithm in the deep reinforcement learning algorithm. The A3C algorithm is a deep reinforcement learning algorithm which combines a use value function and a strategy gradient. The actor part can dynamically change the strategy according to the learned value function. The critic part estimates the current state (action) value function and evaluates the actor's strategy [38]. The basic framework of the A3C algorithm based on the SIN is shown in Figure 5.

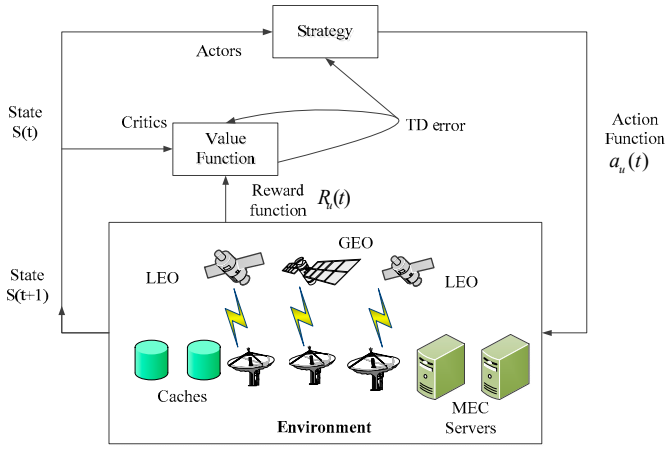

Figure 5. Framework of the Asynchronous Advantage Actor-Critic (A3C) algorithm based on the SIN.

In the A3C algorithm, first of all, we define the learning strategy as ι. The value function $V^{\iota}(s)$ and action value function $Q^{\iota}(s,a)$ are used to judge the learning strategy. The value function $V^{\iota}(s)$ of the current initial state s is defined as

$$V^{\iota}(s) = E^{\iota}[\sum_{k=0}^{\infty} {}^{k} R_u(t+k+1)_i | S_t = s], \tag{27}$$

where $E^{\iota}[*]$ represents mathematical expectations under certain state transition probabilities and learning strategies, $R_u(t)$ represents the reward function, and is the discount factor, $\in [0,1]$. is used to measure the role of reward function in value function. The farther it is away from the current state, the smaller the value of will be.

Each strategy represents a mapping from state to action space, i.e., $a = \iota(s)$. The action value function $Q^{\iota}(s,a)$ is defined as

$$Q^{\iota}(s,a) = E^{\iota}[\sum_{k=0}^{\infty} {}^{k} R_u(t+k+1) | S_t = s, a_u(t) = a]. \tag{28}$$

Actor networks can be divided into three parts. Assuming that the network parameter of the Actor part is Θ, the following results are obtained:

(1) Revenue function: $J(\Theta) = V^{\iota\Theta}(s) = E_{\iota\Theta}[V]$;

(2) Derivation of strategy function: $\nabla_{\Theta}\iota_{\Theta}(s,a) = \iota_{\Theta}(s,a)\nabla_{\Theta}\log\iota_{\Theta}(s,a)$;

(3) Renewal of income gradient through gradient: $\nabla_{\Theta}J(\Theta) = E_{\iota_{\Theta}}(s,a)[\nabla_{\Theta}\log\iota_{\Theta}(s,a)V^{\iota\Theta}(s)]$.

For the Critic part, set the network parameter as Θ_c. When the Actor network and the Critic network are finally determined, $V_{\Theta_c}(s) \approx V^{\iota\Theta}(s)$. The optimal strategy obtained through Actor and Critic networks is the same. Therefore, the gradients of the two should be equal, i.e., $\nabla_{\Theta_c}V_{\Theta_c}(s) = \nabla_{\Theta}\log\iota_{\Theta}(s,a)$.

After the above deduction, we define the loss function as $\varepsilon = E_\iota[(V^{\iota\Theta}(s) - V_{\Theta_c}(s))^2]$. When the loss function is minimized, its minimum value is obtained when the derivative is 0. It can be concluded that $\nabla_{\Theta_c}\varepsilon = 0$. Further derivation shows that

$$\begin{aligned} &E_\iota[(V^{\iota\Theta}(s) - V_{\Theta_c}(s))\nabla_{\Theta_c} V_{\Theta_c}(s)] = 0 \\ &E_\iota[(V^{\iota\Theta}(s) - V_{\Theta_c}(s))\nabla_{\Theta} \log \iota_\Theta(s,a)] = 0 \\ &E_\iota[V^{\iota\Theta}(s)\nabla_{\Theta} \log \iota_\Theta(s,a)] = E_\iota[V_{\Theta_c}(s)\nabla_{\Theta} \log \iota_\Theta(s,a)] \end{aligned} \tag{29}$$

Therefore, the gradient of the income function $J(\Theta)$ is

$$\nabla_\Theta J(\Theta) = E_{\iota_\Theta}[\nabla_\Theta \log \iota_\Theta(s,a) V_{\Theta_c}(s)]. \tag{30}$$

It is known that the network parameter of Actor part is Θ and that of the Critic part is Θ_c. Since there are multiple threads in the A3C algorithm, we have two parameters in the thread: Θ' and ${\Theta_c}'$. Set the global counter $T = 0$; thus, each thread has its own counter t. The flow chart of the A3C algorithm is shown below.

Algorithm: Asynchronous Advantage Actor-Critic

Initialize thread step counter $t \leftarrow 1$
repeat
  Reset gradients: $d\Theta \leftarrow 0$ and $d\Theta \leftarrow 0$ $d\Theta_c \leftarrow 0$
  Synchronize thread-specific parameters $\Theta' = \Theta$ and ${\Theta_c}' = \Theta_c$
  $t_{start} = t$
  Get state S_t
  repeat
    Perform $a_u(t)$ according to policy $\iota(a_u(t)|S_t;\Theta')$
    Receive reward $R_u(t)$ and new state S_{t+1}
    $t \leftarrow t + 1$
    $T \leftarrow T + 1$
  until terminal S_t or $t - t_{start} = t_{\max}$
  $\mathrm{R} = \begin{cases} 0 & \text{forterminal } S_t \\ V(S_t, \Theta_c') & \text{fornon} - \text{terminal } S_t \text{//Bootstrap from last state} \end{cases}$
  for $k \in \{t-1, \ldots, t_{start}\}$ do
    $R \leftarrow R_u(k+1) + R$
    Accumulate gradients wrt Θ': $d\Theta \leftarrow d\Theta + \nabla_{\Theta'} \log \iota(a_k|S_k;\Theta')(R - V(S_k;{\Theta'}_c))$
Accumulate gradients wrt ${\Theta_c}'$: $d\Theta_c \leftarrow d\Theta_c + \partial(R - V(S_k;\Theta_c')^2)/\partial\Theta_c'$
end for
Perform asynchronous update of Θ using $d\Theta$ and of Θ_c using $d\Theta_c$
Until $T > T_{\max}$

(1) Thread counters are initialized to $t = 1$. The network parameters Θ and Θ_c are used to initialize the parameters Θ' and Θ_c' in the thread.

(2) Iterate sequentially until the maximum number of executions $t_{\max}$ is reached, or other termination states are encountered. In successive iterations, the action $a_u(t)$ is obtained by using the strategy function $\iota(a_u(t)|S_t;\Theta')$. Execute this action to get the next state $S(t+1)$ and the corresponding reward value $R_u(t)$. The value function of each state is solved by the Critic network at this time.

$$R = \begin{cases} 0 & \text{In case of termination} \\ V(S_t, \Theta_c') & \text{General situation} \end{cases}$$

Update counters: $t = t + 1$, $T = T + 1$.

(3) In multiple sampling, it may be $t_{\max}$ times, or it may end in advance. The Bellman equation is used to calculate the value function for each sampling result, and the network parameters of Actor and Critic are updated by gradient.

(4) After the number of iterations is reached, the parameters Θ' and $\Theta_c{}'$ in each thread are used to update the network parameters Θ and Θ_c of the whole Actor and Critic parts.

5. Simulation Analysis

5.1. Simulation Parameter Setting

In the experiment, the hardware environment was an Intel Core i7-8750 CPU, with 8 GB of memory and 1 TB of hard disk space. The software environment was Python3.6.1 with Tensorflow1.4.0, MATLAB R2014a [39].

We assumed that there were three GEO data relay satellites, five LEO communication satellites, seven MEC servers, and seven caches. The altitudes of the five LEO satellites were 500 km, 780 km, 1000 km, 1200 km, and 1400 km. The elevation angle between user u and LEO satellite l conforms to Markov chain model. Assuming that the elevation angle is excellent, $w_u^l = 10$, better, $w_u^l = 8$, medium elevation, $w_u^l = 6$, lower elevation, $w_u^l = 4$, and extremely bad, $w_u^l = 2$. We assume that the elevation state transition probability matrix is

$$\kappa = \begin{bmatrix} 0.4\ 0.1\ 0.2\ 0.2\ 0.1 \\ 0.1\ 0.4\ 0.1\ 0.2\ 0.2 \\ 0.2\ 0.1\ 0.4\ 0.1\ 0.2 \\ 0.2\ 0.2\ 0.1\ 0.4\ 0.1 \\ 0.1\ 0.2\ 0.2\ 0.1\ 0.4 \end{bmatrix}. \tag{31}$$

Similarly, when the communication efficiency between user u and satellite l is very excellent, the spectrum utilization ratio is $v_u^l(t) = 10$, better, $v_u^l(t) = 8$, medium condition, $v_u^l(t) = 5$, lower condition, $v_u^l(t) = 1$, and extremely bad, $v_u^l(t) = 0.2$. Its state transition probability matrix is

$$l = \begin{bmatrix} 0.5\ 0.1\ 0.05\ 0.15\ 0.3 \\ 0.3\ 0.5\ 0.1\ 0.05\ 0.15 \\ 0.15\ 0.3\ 0.5\ 0.1\ 0.05 \\ 0.05\ 0.15\ 0.3\ 0.5\ 0.1 \\ 0.1\ 0.05\ 0.15\ 0.3\ 0.5 \end{bmatrix}. \tag{32}$$

Assuming that there is a space task, whether it needs a GEO relay satellite transmission conforms to a Markov chain model, and its state transition probability matrix is

$$\diamond = \begin{bmatrix} 0.3\ 0.7 \\ 0.7\ 0.3 \end{bmatrix}. \tag{33}$$

The cache state of the space task conforms to the Markov chain model, and its state transition probability matrix is

$$\Phi = \begin{bmatrix} 0.6\ 0.4 \\ 0.4\ 0.6 \end{bmatrix}. \tag{34}$$

For MEC servers, when the computing state is excellent, the computing rate is $\Xi_u^m(t) = 50$, better, $\Xi_u^m(t) = 30$, medium condition, $\Xi_u^m(t) = 10$, lower condition, $\Xi_u^m(t) = 3$, and extremely bad, $\Xi_u^m(t) = 0.5$. Its state transition probability matrix is

$$E = \begin{bmatrix} 0.5\ 0.15\ 0.05\ 0.25\ 0.05 \\ 0.05\ 0.5\ 0.15\ 0.05\ 0.25 \\ 0.25\ 0.05\ 0.5\ 0.15\ 0.05 \\ 0.05\ 0.25\ 0.05\ 0.5\ 0.15 \\ 0.15\ 0.05\ 0.25\ 0.05\ 0.5 \end{bmatrix}. \tag{35}$$

The remaining parameters in the simulation are shown in Table 2.

Table 2. Simulation parameter setting. LEO—low Earth orbit; GEO—geostationary orbit; CPU—central processing unit.

Parameters	Values	Descriptions
B_u^l	6 MHz	Bandwidth allocated by LEO satellite l to user u
B_l^{lg}	6 MHz	Bandwidth allocated by GEO satellite lg to user l
δ_l	2 units/MHz	Payment price using LEO spectrum resources
δ_{lg}	2 units/MHz	Payment price using GEO spectrum resources
ς_c	4 units/Mbits	Payment price using caching resources
η_m	1 unit/J	Payment price using computing resources
τ_u	15 units/Mbps	The unit transmission fee charged to the user
κ_u	10 units/Mbps	The unit caching fee charged to the user
ϕ_u	5 units/Mbps	The unit computing fee charged to the user
$\theta_{u,\max}^l$	$\pi/2$	Maximum elevation between user u and satellite l
n_u	6 Mcycles	Number of cycles a CPU takes to complete each space task
e_m	1 J	The energy consumed by the CPU in one lap
o_u	3 Mbits	Task content

In this experiment, we simulated the expected benefits of unit resources in the following six situations as follows:

(1) Unified consideration of LEO satellite elevation state, communication link state, GEO data relay satellite transmission state, caching state, and computing state, expressed as A3C-based all scheme.

(2) Unified consideration of GEO data relay satellite transmission state, caching state, and computing state, regardless of LEO satellite elevation state and communication link state, expressed as A3C-based without coverage communication scheme.

(3) Unified consideration of LEO satellite elevation state, communication link state, caching status, and computing state, regardless of GEO data relay satellite transmission state, expressed as A3C-based without GEO communication.

(4) Unified consideration of LEO satellite elevation state, communication link state, GEO data relay satellite transmission state, and computing state, regardless of caching state, expressed as A3C-based without caching scheme.

(5) Unified consideration of LEO satellite elevation state, communication link state, GEO data relay satellite transmission state, and caching state, regardless of computing state, expressed as A3C-based without computing scheme.

(6) Direct allocation of resources under static network conditions, expressed as A3C-based no scheme [40].

5.2. Simulation Result

The simulation results in this paper are discussed below.

Figure 6 shows the convergence performance under different schemes. From the simulation, we can see that, at the beginning of deep reinforcement learning, the expected benefit per unit resource is low. With the increase of training times, the expected benefit of unit resources tends to be stable. The proposed A3C-based all scheme takes into account the coverage area of the LEO satellite, the communication link state between users and the LEO satellite, the transmission state of the GEO data relay satellite, the caching state of caches, and the computing state of the MEC server, which has better resource utilization efficiency.

Figure 7 shows that with the increase in elevation angles of users and LEO satellites, the expected benefits per unit resource of the SIN increase gradually. The proposed A3C-based all scheme takes into account the coverage area of the LEO satellite, the communication link state between users and the

LEO satellite, the transmission state of the GEO data relay satellite, the caching state of caches, and the computing state of the MEC server, which has better resource utilization efficiency.

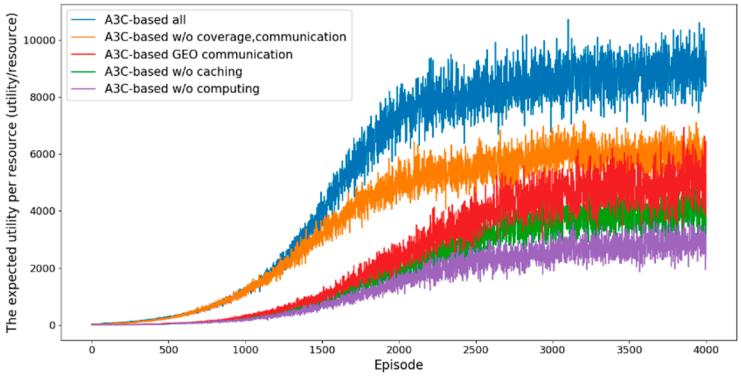

Figure 6. Convergence performance under different schemes.

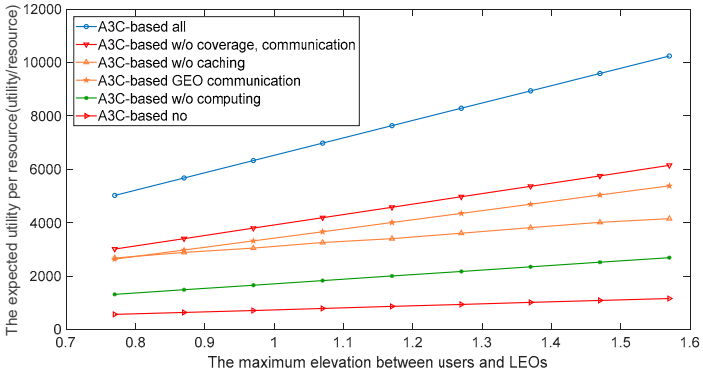

Figure 7. Expected benefits of unit resources under different elevation angles.

Figure 8 shows that, with the increase of the task content, the cost of caching charged to users increases gradually; thus, the expected benefit of unit resources of the SIN decreases gradually. The proposed A3C-based all scheme takes into account the coverage area of the LEO satellite, the communication link state between users and the LEO satellite, the transmission state of the GEO data relay satellite, the caching state of caches, and the computing state of the MEC server, which can achieve better expected benefits per unit resource.

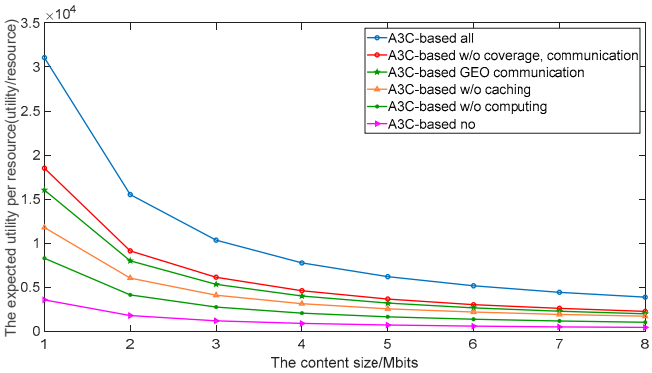

Figure 8. Expected benefits of unit resources under different task content.

Figure 9 shows the relationship between the unit charging price for using transmission resources and the expected benefit of the unit resource. With the increase of the unit charging price for using transmission resources, the expected benefit of the unit resource of the SIN increases gradually.

The scheme of A3C-based all takes into account the coverage area of the LEO satellite, the state of communication link between users and the LEO satellite, the transmission state of the GEO data relay satellite, the caching state of caches, and the computing state of the MEC server. It effectively improves the efficiency of unit resource utilization, and has more advantages than other schemes.

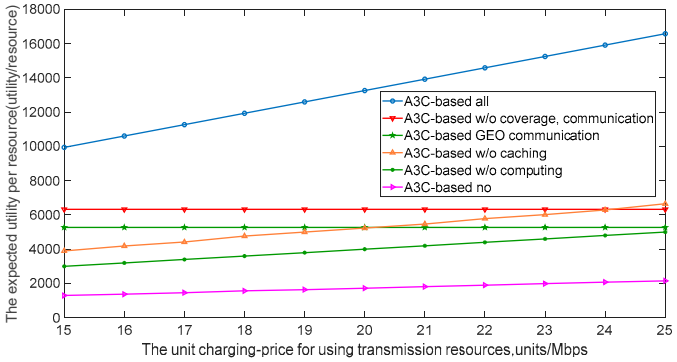

Figure 9. The relationship between the unit charging price for using transmission resources and the expected benefit of unit resources.

Figure 10 shows the relationship between the unit charging price for using caching resources and the expected benefit of the unit resource. With the increase of the unit charging price for using caching resources, the expected benefit of the unit resource of the SIN increases gradually. The scheme of A3C-based all takes into account the coverage area of the LEO satellite, the state of communication link between users and the LEO satellite, the transmission state of the GEO data relay satellite, the caching state of caches, and the computing state of the MEC server. It effectively improves the efficiency of unit resource utilization, and has more advantages than other schemes.

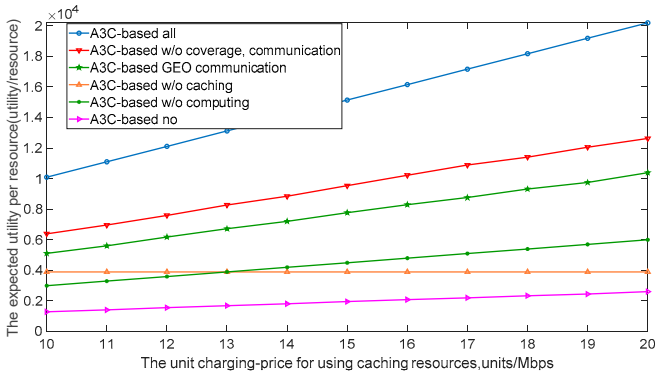

Figure 10. The relationship between the unit charging price for using caching resources and the expected benefit of unit resources.

Figure 11 shows the relationship between the unit charging price for using computing resources and the expected benefit of the unit resource. With the increase of the unit charging price for using caching resources, the expected benefit of the unit resource of the SIN increases gradually. The proposed A3C-based all scheme takes into account the coverage area of the LEO satellite, the communication link state between users and the LEO satellite, the transmission state of the GEO data relay satellite, the caching state of caches, and the computing state of the MEC server, which has better resource utilization efficiency.

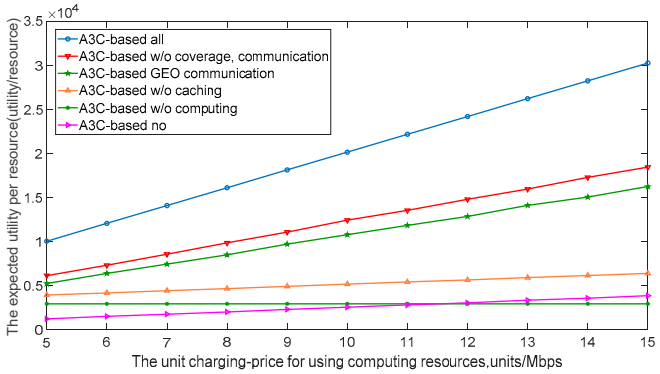

Figure 11. The relationship between the unit charging price for using computing resources and the expected benefit of unit resources.

6. Conclusions

In this paper, in order to improve the resource management and utilization efficiency of the SIN, firstly, based on the core idea of SDN, a hierarchical and domain-controlled SIN architecture was established. The overall networking architecture and network control architecture were designed. On this basis, the transmission resources, caching resources, and computing resources of the SIN were managed in a unified way. Next, the satellite coverage and transmission model, communication link model, caching model, and computing model of the SIN were modeled and analyzed. Finally, the A3C algorithm of deep reinforcement learning was introduced to model and simulate the multi-dimensional resource allocation problem of the SIN. The simulation results show that the proposed scheme can effectively improve the expected benefits of unit resources and the utilization efficiency of the SIN resources. In this paper, LEO communication satellites and several GEO data relay satellites were taken as examples for analysis. However, the SIN is a huge system. In practical applications, the scheduling of remote-sensing satellites, navigation satellites, and other resources may have different situations, which need specific analysis. Furthermore, in a follow-up study, we will further analyze the other SIN resources such as energy resources and sensor resources.

Author Contributions: Conceptualization, X.M. and L.W.; Methodology, X.M. and L.W.; Software, X.M. and S.Y.; Validation, X.M.; Formal Analysis, X.M.; Investigation, X.M.; Resources, X.M. and S.Y.; Data Curation, X.M.; Writing-Original Draft Preparation, X.M.; Writing-Review & Editing, X.M.; Visualization, X.M.; Supervision, L.W.

Funding: China Equipment Named Research Funded Project with grant number 6142010010301.

Acknowledgments: The authors would like to thank Chao Qiu of Beijing University of Posts and Telecommunications for her guidance on the ideas for the text.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, W. *Topological Control Theory and Method of Space Information Network*; PLA University Science and Technology: Nanjing, China, 2016; pp. 1–5.
2. Wang, Y.; Sheng, M.; Zhuang, W.; Zhang, S.; Zhang, N.; Liu, R. Multi-Resource Coordinate Scheduling for Earth Observation in Space Information Networks. *IEEE J. Sel. Areas Commun.* **2018**, *36*, 268–279. [CrossRef]
3. National Natural Science Foundation. The Program Guidance of the Basic Theory and Key Technology Research of Space Information Network in 2016. Available online: http://www.nsfc.gov.cn/publish/portal0/tab38/info51946.htm (accessed on 25 March 2016).
4. Yu, Q.Y.; Meng, W.X.; Yang, M.C.; Zheng, L.M.; Zhang, Z.Z. Virtual multi-beamforming for distributed satellite clusters in space information networks. *IEEE Wirel. Commun.* **2016**, *23*, 95–101. [CrossRef]
5. Li, D.R.; Shen, X.; Gong, J.Y.; Zhang, J.; Lu, J.H. On construction of China's space information network. *Wuhan Univ. Inf. Sci. Ed.* **2015**, *40*, 711–715. [CrossRef]

6. Cui, L.; Yu, F.R.; Yan, Q. When big data meets software-defined networking: SDN for big data and big data for SDN. *IEEE Netw.* **2016**, *30*, 58–65. [CrossRef]
7. Li, T.X.; Zhou, H.C.; Xu, Q. SAT-FLOW: Multi-Strategy Flow Table Management for Software Defined Satellite Networks. *IEEE Access* **2017**, *5*, 14952–14965. [CrossRef]
8. Gardikis, G.; Koumaras, H.; Sakkas, C.; Koumaras, V. Towards SDN/NFV-enabled satellite networks. *Telecommun. Syst.* **2017**, *66*, 1–14. [CrossRef]
9. Liu, Q.; Zhai, J.W.; Zhang, Z.Z.; Zhong, S.; Zhou, Q.; Zhang, P. A Survey on Deep Reinforcement Learing. *Chin. J. Comp.* **2018**, *1*, 1–27. [CrossRef]
10. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef]
11. Jennings, E.; Heckman, D. Performance Characterization of Space Communications and Navigation (SCaN) Network by Simulation. In Proceedings of the IEEE Aerospace Conference, Big Sky, MT, USA, 1–8 March 2008; pp. 1–9. [CrossRef]
12. Vanderpoorten, J.; Cohen, J.; Moody, J.; Cornell, C.; Streland, A.; Breese, S. Transformational Satellite Communications System (TSAT) lessons learned: Perspectives from TSAT program leaders. In Proceedings of the 2012 IEEE Military Communications Conference, Orlando, FL, USA, 29 October–1 November 2012; pp. 1–6. [CrossRef]
13. Sesena, J.; Alfaro, A.; Munoz, S. Regulatory environment for the successful ISICOM development. In Proceedings of the 2009 International Workshop on Satellite and Space Communications, Tuscany, Italy, 9–11 September 2009; pp. 109–112. [CrossRef]
14. Axford, R.; Short, S.; Shchupak, P.; Muhammad, N. Wideband Global SATCOM (WGS) earth terminal interoperability demonstrations. In Proceedings of the 2012 IEEE Military Communications Conference, San Diego, CA, USA, 16–19 November 2008; pp. 1–6. [CrossRef]
15. Schroth, K.; Burkhardt, N.; Che, T.S.; Pisano, D. IP networking over the AEHF MILSATCOM system. In Proceedings of the 2012 IEEE Military Communications Conference, Orlando, FL, USA, 29 October–1 November 2012; pp. 1–6. [CrossRef]
16. Adinolfi, M.; Cesta, A. Heuristic scheduling of the DRS communication system. *Eng. Appl. Artif. Intell.* **1995**, *8*, 147–156. [CrossRef]
17. Rojanasoonthon, S.; Bard, J.F.; Reddy, S.D. Algorithms for parallel machine scheduling: A case study of the tracking and data relay satellite system. *J. Oper. Res. Soc.* **2003**, *54*, 806–821. [CrossRef]
18. Gu, Z.S. *Research on the Relay Satellite Dynamic Scheduling Problem Modeling and Optimizational Technology*; National University of Defense Technology: Changsha, China, 2008; pp. 11–26. [CrossRef]
19. Bertaux, L.; Medjiah, S.; Berthou, P.; Abdellatif, S.; Hakiri, A.; Gelard, P.; Planchou, F.; Bruyere, M. Software defined networking and virtualization for broadband satellite networks. *IEEE Commun. Mag.* **2015**, *53*, 54–60. [CrossRef]
20. Ferrús, R.; Koumaras, H.; Sallent, O.; Agapiou, G.; Rasheed, T.; Kourtis, M.-A.; Boustie, C.; Gélard, P.; Ahmed, T. SDN/NFV-enabled satellite communications networks: Opportunities, scenarios and challenges. *Phys. Commun.* **2016**, *18*, 95–112. [CrossRef]
21. Gopal, R.; Ravishankar, C. Software Defined Satellite Networks. In Proceedings of the Aiaa International Communications Satellite Systems Conference, San Diego, CA, USA, 24–27 September 2013. [CrossRef]
22. Yu, X.; Lei, W.M.; Song, L. A framework of SDN-based satellits on-board switching networks. *J. PLA Univ. Sci. Tech. (Nat. Sci. Ed.)* **2017**, *18*, 224–230. [CrossRef]
23. Zhu, S.Y. *Research on Routing Algorithm of Space Network Based on SDN*; Harbin Institute of Technology: Harbin, China, 2017; pp. 1–19.
24. Tian, R.; Yu, X.S.; Zhao, Y.L.; Wang, W.Z.; Li, Y.J.; Wang, C.F.; Zhang, J. Multi-path Carrying Strategy in SDN-based Space Information Networks. *Radio Eng.* **2016**, *46*, 63–67. [CrossRef]
25. Tian, R. *Research on Control Protocol and Routing Algorithms of Software Defined Space-Terrestrial Network*; Beijing University of Posts and Telecommunications: Beijing, China, 2017; pp. 9–16.
26. Zhang, S.M.; Zou, F.M. Survey on software defined network research. *Appl. Res. Comput.* **2013**, *30*, 2246–2251. [CrossRef]
27. Nguyen, X.N.; Saucez, D.; Barakat, C. Rules Placement Problem in OpenFlow Networks: A Survey. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 1273–1286. [CrossRef]

28. Zhang, Q.; Li, M.; Deng, Y. Measure the structure similarity of nodes in complex networks based on relative entropy. *Phys. A Stat. Mech. Appl.* **2018**, *491*, 749–763. [CrossRef]
29. Yang, B.; He, F.; Jin, J.; Xu, G.H. Analysis of Coverage Time and Handoff Number on LEO Satellite Comunication Systems. *J. Electron. Inf. Technol.* **2014**, *36*, 804–809. [CrossRef]
30. Deng, B.; Jiang, C.; Kuang, L.; Guo, S.; Lu, J.; Zhao, S. Two-Phase Task Scheduling in Data Relay Satellite Systems. *IEEE Trans. Veh. Technol.* **2018**, *67*, 1782–1793. [CrossRef]
31. Gomaa, H.; Messier, G.G.; Williamson, C.; Davies, R. Estimating Instantaneous Cache Hit Ratio Using Markov Chain Analysis. *IEEE/ACM Trans. Netw.* **2013**, *21*, 1472–1483. [CrossRef]
32. Breslau, L.; Cao, P.; Fan, L.; Phillips, G.; Shenker, S. Web caching and Zipf-like distributions: Evidence and implications. *Proc. IEEE INFOCOM* **1999**, *1*, 126–134. [CrossRef]
33. Li, H.Q. *Hardware Implementation of LEO Satellite Channel Characteristic Emulation*; Harbin Institute of Technology: Harbin, China, 2008; pp. 12–15.
34. Theofanis, X.; Psannis, K.E. Caching Hit Probability and Compressive Sensing Perspective for Mobile Cellular Networks. *Simul. Model. Pract. Theory* **2018**, *87*, 92–98. [CrossRef]
35. Daniel, G.; Gerson, S.; Jordi, C. Advanced prefetching and caching of models with PrefetchML. *Softw. Syst. Model.* **2018**, 1–22. [CrossRef]
36. Zhou, Y.; Yu, F.R.; Chen, J.; Kuo, Y. Resource Allocation for Information-Centric Virtualized Heterogeneous Networks with In-Network Caching and Mobile Edge Computing. *IEEE Trans Veh. Technol.* **2017**, *66*, 11339–11351. [CrossRef]
37. He, Y.; Zhao, N.; Yin, H.X. Integrated Networking, Caching and Computing for Connected Vehicles: A Deep Reinforcement Learning Approach. *IEEE Trans. Veh. Technol.* **2018**, *67*, 44–55. [CrossRef]
38. Helma, C.; Cramer, T.; Kramer, S.; Raedt, L.D. Data Mining and Machine Learning Techniques for the Identification of Mutagenicity Inducing Substructures and Structure Activity Relationships of Noncongeneric Compounds. *J. Chem. Inf. Comput. Sci.* **2018**, *35*, 1402–1411. [CrossRef]
39. Jiang, S.W.; Guo, K.K.; Liao, J.; Zheng, G.A. Solving Fourier ptychographic imaging problems via neural network modeling and TensorFlow. *Biomed. Opt. Express* **2018**, *9*, 3306–3319. [CrossRef]
40. Ying, H.; Cheng, C.L.; Richard, Y.; Zhu, H. Trust-based Social Networks with Computing, Caching and Communications: A Deep Reinforcement Learning Approach. *IEEE Trans. Netw. Sci. Eng.* **2018**, 1–14. [CrossRef]