Intelligent Automation & Soft Computing
DOI:10.32604/iasc.2022.022536
Article

Tech Science Press

Deep Reinforcement Learning-Based Long Short-Term Memory for Satellite IoT Channel Allocation

S. Lakshmi Durga[1], Ch. Rajeshwari[1], Khalid Hamed Allehaibi[2], Nishu Gupta[3,*], Nasser Nammas Albaqami[4], Isha Bharti[5] and Ahmad Hoirul Basori[6]

[1]Electronics and Communication Engineering Department, Vaagdevi College of Engineering, Warangal, 506005, India
[2]Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[3]Electronics and Communication Engineering Department, Chandigarh University, Mohali, 160036, India
[4]Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[5]Senior Business Analyst & Solution Architect, SAP Technology & Innovation, Capgemini America Inc., 75039, USA
[6]Department of Information Technology, Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, 21589, Saudi Arabia
*Corresponding Author: Nishu Gupta. Email: nishugupta@ieee.org
Received: 10 August 2021; Accepted: 26 September 2021

Abstract: In recent years, the demand for smart wireless communication technology has increased tremendously, and it urges to extend internet services globally with high reliability, less cost and minimal delay. In this connection, low earth orbit (LEO) satellites have played prominent role by reducing the terrestrial infrastructure facilities and providing global coverage all over the earth with the help of satellite internet of things (SIoT). LEO **satellites** provide wide coverage area to dynamically accessing network with limited resources. Presently, most resource allocation schemes are designed only for geostationary earth orbit (GEO) satellites. For LEO satellites, resource allocation is challenging due to limited availability of resources. Moreover, due to uneven distribution of users on the ground, the satellite remains unaware of the users in each beam and therefore cannot adapt to changing state of users among the beams. In this paper, long short-term memory (LSTM) neural network has been implemented for efficient allocation of channels with the help of deep reinforcement learning (DRL) model. We name this model as DRL-LSTM scheme. Depending on the pool of resources available to the satellite, a channel allocation method based on the user density in each beam is designed. To make the satellite aware of the number of users in each beam, previous information related to the user density is provided to LSTM. It stores the information and allocates channels depending upon the requirement. Extensive simulations have been carried out which have shown that the DRL-LSTM scheme performs better as compared to the traditional and recently proposed schemes.

Keywords: Artificial intelligence; channel allocation; deep reinforcement learning; LSTM; satellite internet of things; supervised training

1 Introduction

Satellite internet of things (SIoT) plays a dominant role in communication system for future wireless network [1]. Satellite network provides numerous advantages including wide-coverage area, addressing interruption problems in communication systems, etc. [2,3]. Integration of wireless communication with internet of things (IoT) alleviates the number of transmissions and provides greater efficiency [4]. According to the ABI report [5], by the year 2024 there will be an increase of 24 million IoT connected devices globally. Due to lack of terrestrial infrastructure, the demand for SIoT have to be increased to drive the end-devices at remote sites [6]. Moreover, satellite network plays a significant role in IoT as terrestrial networks are prone to natural attacks like earthquakes and tsunamis. To resolve such issues and to provide communication ability in remote areas, especially for the internet of remote things (IoRT) which provides network access to IoT devices in rural and urban areas, low earth orbit (LEO) satellite constellation has been introduced. Apart from LEO satellites there are geostationary earth orbit (GEO) satellite also, but LEO satellites offer better features such as high-speed internet services for the customers. Due to its low latency, the satellite transmits and receives the data quickly [7,8]. LEO satellite network provides latitude of 160 km to 2000 km. To minimize end-to-end propagation delay, for better communication, wide coverage area and assured quality of services; many LEO satellite based commercial networks such as '*SpaceX*', '*OneWeb*', '*LeoSat*' and '*Iridium*' have come into existence [9,10]. LEO satellite based IoT services possess the potential of long-distance communication in remote areas. Also, the sensor data acquisition is an important supplement of ground based IoT that transmits data periodically. It can help in saving power along with extended battery life till the existence of satellite [11]. To provide global connectivity in dynamic environment and real-time communication '*Iridium*' has designed a mesh architecture which comprises 66 LEO satellites with the collaboration of '*long range*' (LoRa). It is one of the dominant modulation technologies of low power wireless area network (LPWAN) developed by *Semtech*, an American company with IoT platform provider stream technology. Deep reinforcement learning (DRL) is a promising technique to solve many optimization problems for network models. To avoid congestion and unbalanced traffic distribution in LEO satellite constellation, DRL technique with a scheme of explicit load balancing and link handling is introduced [12]. Link handling between satellite and IOT devices can operate in two modes, viz. indirect access and direct access. In direct access mode, users directly communicate with satellite through sensors and actuators, whereas in indirect access mode these sensors and actuators communicate with satellite through a communication gateway [13]. To provide communication between satellite and end devices there are some existing protocols. *LoRa* is one such protocol which is used in physical layer [14]. It has the capability of providing long distance communication with low power and also provides high robustness against noise from multipath propagation inferences. *LoRa* transmits in several sub-gigahertz frequencies for many applications. At industrial scientific and medical (ISM) band, *LoRa* has adopted a chirp spread spectrum modulation in physical layer, and it also uses ultra-narrow band technique.

LEO satellite constellations provide less than 100 ms of round-trip time (RTT) as compared with GEO satellite. Resource allocation is a serious issue in LEO, but it provides better signal strength as compared to GEO satellite. As the LEO satellite size is smaller than GEO, it has limited energy and resources [15]. Therefore, serving these limited resources with over the coverage area is a challenging task .To overcome this challenge, we propose DRL based long short-term memory (LSTM) model to predict the traffic values and to allocate resources with low power energy consumption. LSTM is a novel recurrent neural network (RNN) to store relevant information, which builds relationship between the present and past information. LSTM neural network is commonly used in prediction of time series information. DRL solves the decision-making problems very efficiently by combining reinforcement learning (RL) with deep learning (DL). It displays significant improvement in resource allocation [16,17]. Based on these techniques, one of the significant achievements in the proposed DRL-LSTM scheme is that the

satisfaction rate is higher as compared to all the other schemes. This is due to the advantage of training LSTM with previous scenarios.

1.1 Basic Mechanism of DRL

Deep reinforcement learning (DRL) solves many real-world problems like controlling the traffic signals [18]. In DRL, the agent learns the optimal policy by interacting with the environment for mapping states to corresponding actions. Here, the agent must be capable of sensing the present input state and its corresponding actions. With the rapid development of social networking sites, privacy and storage have become a serious issue as the transfer of a huge amount of data to the cloud has become difficult. Solutions to this problem are provided by DRL [19,20] integration of DRL with deep neural network (DNN) results in effective functioning of DRL algorithms [21]. This is highly successful in the fields of robotics, finance, healthcare, videogames, etc. Many unsolved problems have been solved by this model [22]. To extract high dimensional observation features in RL, we implement DRL by applying a DNN to it. Presently, the demand for SIoT devices with direct access has drastically increased with the requirement of low cost and long battery life. Above it, DRL has grasped much attention with advancements in algorithm for resource and channel allocation in satellites. To maximize the channel utilization and throughput, TDMA scheme is introduced [23].

1.2 Contribution of the Proposed Work

1. It solves the problem of resource allocation in complex environment of LEO satellite IoT systems. We have proposed a DRL model with LSTM neural network (DRL-LSTM) which stores information of environmental load scenarios of previous user for analyzing the traffic demand of the present state.
2. Markov Decision Process (MDP) is considered that takes action based on the cell state of LSTM and receives reward at each episode.
3. It solves the existing channel allocation problem in LEO satellite, when it is in complex dynamic environment. For that, we have considered entire footprint of the satellite as square matrix where each square matrix is divided into sub-matrices consisting of the information of user requests.
4. It optimizes the problem of limited resources and large user requests. The resource allocation (i.e., power, channel) is implemented using LSTM.

1.3 Organization of the Paper

The rest of the paper is organized as follows. Section 2 discusses the related works based on LEO SIoT, characteristics and protocols used in LEO SIoT. Section 3 defines the system model for LEO satellite communication. In Section 4, the proposed DRL-LSTM scheme is discussed. Section 5 displays the results and discussion. Finally, the conclusions of the proposed work is reported in Section 6.

2 Related Works

Conventionally, a lot of research is going on the resource allocation that minimizes the power and saves energy in SIoT. This section is divided into three sub-sections in order to categorically present the literature survey and identify the research gaps in this promising area of research.

2.1 Literature Survey and Research Gaps Based on LEO SIoT

LEO satellite has altitude between 160 km to 2000 km from the surface of the earth which takes an orbital period of 128 min or less. LEO satellites require less power for transmission, higher bandwidth, low latency and higher altitude for communication. LEO satellites are backbone for communication in terrestrial network including desert, deep sea, remote areas, etc. Two largest LEO satellite constellations

are '*Iridium*' and '*Global Star*', which launched 66 and 24 satellites, respectively. *Semtech*, finished testing the first phase of satellite communication connected to *LoRa* technology [24]. **'***Orbocomm'*, an American company provides commercial, low-orbit satellite that uses bidirectional communication for data transmission. Each satellite has a coverage radius of 5100 km. '*Orbocomm*' satellite has multiple applications in positioning of vehicles, transmission and retrieval of mails, remote data collection, etc. [25].

In wireless communication systems, reliability is still a challenging issue that has to be overcome by SIoT, which is believed to provide high security and protection. Coverage of a satellite depends on its beam width. This beam width with multicasting makes SIoT easier [26]. '*SpaceX*' has planned nearly 12,000 satellites approved by federal communication commission (FCC). Meanwhile, '*OneWeb*' hopes to launch a rocket carrying 36 satellites by the mid of 2022 for network communication of 648 units. The computational technologies to overcome the overlapping between spot beams with handover technique has been discussed [27] which is named as *Handover with Queue*. Authors [28] proposed a method using non-orthogonal multiple access (NOMA) scheme for SIoT downlink system to overcome long-term network utility problems. Author in [29] proposed a scheme to overcome the problem of spectrum shortage in terrestrial-based IoT network. Authors [30] introduced a network utility maximization resource allocation for NOMA in SIOT.

2.2 Literature Survey and Research Gaps Based on Characteristics of LEO SIoT

There are several reasons for satellite communication to integrate with IoT technology. LEO satellite based IoT has the ability to provide wide-coverage area, low power consumption and narrow band technology [11]. Authors [6] discuss delay sensitive application (DSA) used for disaster recovery. To provide automatic storage and forward data in communication network, delay tolerant application (DTA) is presented [6]. To provide channel allocation for users depending on traffic demand, demand assigned multiple access method (DAMA) is presented [31]. To avoid collision between two users in communication network and to access contention-free data transmission, enhanced aloha protocol (EAP) method is discussed [32].

Narrow band IoT (NB-IoT) technology used for data transmission is developed by 3GPP (3^{rd} generation partnership project). It is a sub-class of long-term evolution (LTE) used for energy efficiency [33]. NB-IoT is used in broadband communication with bandwidth of 180 kHz with less energy consumption. It operates in three modes: In-band, Guard-band LTE, Standalone. In-band mode uses the frequencies within LTE carriers, Guard-band LTE mode uses the unused frequencies out of LTE band, and Standalone mode currently uses GSM frequency band.

For satellite based IoT, fixed channel allocation scheme cannot meet the requirements of transmission demands. To acquire these demands and requirements of the users, some protocols have been proposed to improve energy efficiency. Development of DRL based dynamic channel allocation (DRL-DCA) algorithm is proposed [34] which displays the increased service blocking probability. Accordingly, to alleviate the power consumption and augment the satellite lifetime, a green satellite routing scheme is proposed by putting nodes in sleep mode according to traffic availability using geographic hierarchical routing protocol (GHRP) [15].

2.3 Literature Survey and Research Gaps Based on Protocols Used in LEO SIoT

In SIoT, physical layer mainly focuses to improve upon the channel efficiency. In order to have high power efficiency and very high immunity to interference, long range low power wireless area network (LoRa LPWAN) protocol has been proposed [35]. It has a battery life of >10 years. It uses the license free sub-GHZ radio frequency such as 923 MHz and the communication frequencies. Power and data rate are also managed. *LoRa* uses chirp spread spectrum (CSS) modulation where the signal power is spread

with a spreading code. *LoRa* allows device interconnectivity and rapid formation of IoT applications. Moreover, it keeps up the connection with the devices which are in motion without any stress on power with end-to-end encryption and mutual authentication [36]. For better transmission efficiency, the link layer medium access protocol (MAC) commonly used is TDMA. Here, the channels are assigned with fixed time slots and every channel is operated at the beginning of the time slot. However, this assignment can be changed, based upon the load capacity. This means that the heavier load is assigned a bigger time slot. The main advantage of TDMA is its low power consumption [37].

In LEO satellite constellations, the network layer is used to minimize energy consumption and save power, when it is connected with nodes on the ground in a complex environment. In order to extend battery lifetime, energy efficient routing protocols are implemented by swapping nodes into sleep mode. It distributes the resources based on traffic demand with some green satellite routing schemes such as minimum battery cost routing (MBCR) and minimum total transmission power (MTTP), low energy adaptive clustering hierarchy (LEACH), etc. [38]. Authors discuss a cluster-based routing scheme named high-energy efficient distributed routing scheme (HEED) to improve battery life [39]. Furthermore, LPWAN device is used for managing energy consumption and providing long battery life, low-bit rate, and it provides bandwidth of 12 kHz [40].

3 System Model for LEO Satellite Communication

We consider a modern LEO satellite whose altitude is around 1200 km. A LEO satellite covers several thousand kilometers on the surface of the earth. The ground area that is covered by the microwave radiation emitted from the satellite dish is known as the footprint of the satellite. This footprint depends on the location of the satellite, size of the beam and the distance of the satellite from the earth's surface. Whenever the beam from the satellite falls on the ground, each beam covers a region as shown in Fig. 1. The entire footprint is covered with small cells. Each cell contains user information. In some cells, there are a large number of users; and in some cells, there is only one user.

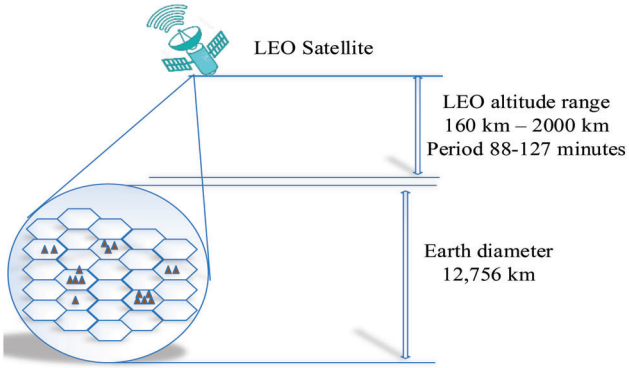

Figure 1: Geographical footprint of a LEO satellite

The maximum duration that a satellite occupies on the ground is known as the visibility period of the satellite as shown in Fig. 2. Let us assume that the coverage area of each beam is X.

There arise three scenarios. Firstly, let us assume that the area under one beam has 'n' number of users. If the beam has the capacity to serve these n users, then it is efficiently used by all the users belonging to that particular region. In the second case, if the area under the beam has '$n-k$' users, then the spectrum remains unused by 'k' number of users. This is because of the lack of users in that particular region. In the third case, if there are '$n+k$' user requests to the beam but the beam can serve only n users, then the spectrum falls short by

k number of users. The above three cases show that the spectrum is not efficiently used when the requested channels by the users are higher or lower than the beam capacity. This problem is due to the fact that the satellites are unaware of the number of users in each region. To solve this problem, we consider the entire footprint of the satellite into a square block which is sub-divided into mini-square blocks (i.e., blk_{11}, blk_{12}—blk_{mn}) as shown in Fig. 3.

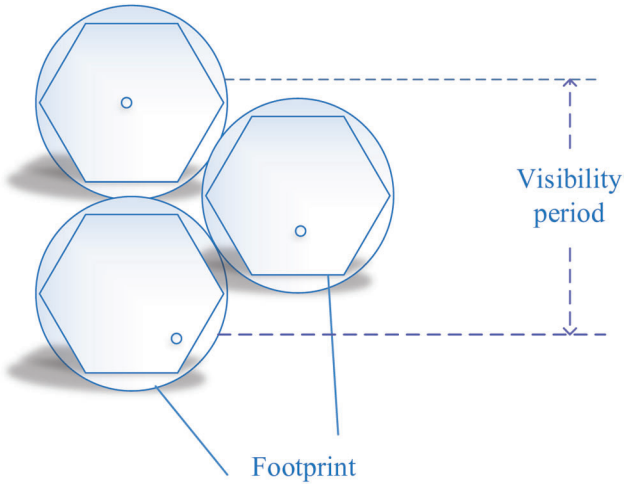

Figure 2: Visibility period of a LEO satellite

blk_{11}	blk_{12}	—	blk_{1n}
blk_{21}	blk_{22}	—	blk_{2n}
		—	
blk_{m1}	blk_{m2}	—	blk_{mn}

Figure 3: Footprint divided into square matrix

Each block has the information about the number of users in the form of matrix (image) of *M* rows and *N* columns. Each square matrix has *n* number of users. Whenever the satellite beam falls on the ground, the area which is exactly under the satellite (i.e., the center region) requires less energy compared to the corners of the footprint as they are far-off from the satellite. During the movement of the satellite, the coverage area is splitted into square blocks and each block is considered as a region. The time duration of the satellite that covers the region at each time step before moving to the next region is denoted by T_b. Moreover, as LEO satellite has limited resources, serving all the users efficiently is a challenging task because the number of requests is greater than the availability of resources. Hence, it is to be ensured that there lies no unused resource in any region of the footprint as this underutilized resource in one region may be responsible for the lack of resource in another region.

3.1 Problem Definition

Let the satellite create $'N_b^{'}$ beams, i.e., b= (n|n = 1, 2, 3, 4, 5, 6,…N_b) and the total number of channels available is denoted by $'N_c^{'}$ i.e., C= (c|c = 1, 2, 3, 4, ….N_c). Each beam consists of Y number of channels allocated to it. Each cell is allocated to a set of channels. If each cell has 'n' number of users and the channel allocation is based on a fixed time slot, the cell uses the channel within the allocated time. After the completion of the time slot, the channel starts serving another cell and leaves the present cell. In the proposed scenario, time slot division is done based on the traffic demand requests. Whenever the spot beam falls on the square block, it serves the users within that block with the channels allocated to that beam. Each channel serves more than one cell, and each cell has different number of users. Assuming that if channel 1 serves both cell 1 and cell 3, then the user requests in cell 1 is 4 and user requests in cell 3 is 2, that is, they are different. This channel information is represented in Fig. 4.

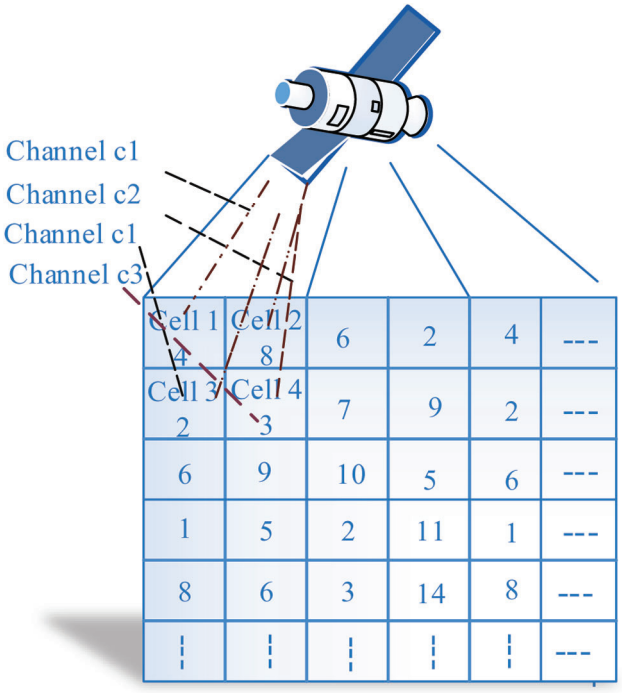

Figure 4: User request on the ground

Each channel can serve only limited number of users within the specified time slot. There are few cases where the number of users in a cell are greater than the beam capacity. In such cases, the users are not served by the channel. Therefore, the channel allocation becomes challenging. We propose DRL-LSTM model for efficient allocation of channel to the beams based on the user requests. Past user information of all the beams in a satellite is given to the LSTM network along with the present user information as input in the form of matrix representation which contains the information of all the user requests belonging to the footprint. In practical, it is not possible for the agent to train as the LSTM neural network. Therefore, the state space is normalized so that it is suitable for the input to the LSTM neural network. Fig. 5 shows the example for normalization of state space. We follow the latest research which assumes that each beam in the footprint of the satellite can serve only 10 users [41]. In Fig. 5, left side matrix represents the number of user requests in the footprint and the right-side matrix represents the normalized matrix. Normalization process is attained by dividing each element in the matrix with maximum beam capacity (i.e., 10 users).

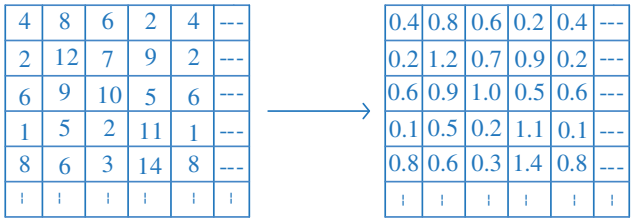

Figure 5: State representation example with normalization

The above matrix represents the user request information of the footprint. To allocate available resources to the entire footprint, several matrix sets are given as input to LSTM neural network at different episodes. This LSTM cell undergoes a training process and allocates resources based on the input given to it. One of the main purposes of the proposed scheme is to serve as much users as possible with minimum power usage. The terms involved in power allocation are: i) total transmitting power of satellite P_{tot}, ii) power allocated to each beam P_b, and iii) power of each channel P_c. Bandwidth of each channel (B_c) is given by

$$B_c = B_{tot}/N_C \tag{1}$$

In the next subsection, the problem is formulated to Markov decision process (MDP).

3.2 Markov Decision Process

Many conventional channel allocations do not have prior knowledge on how to make a decision in a complex environment. For this, *RL* with MDP is designed to solve problems in a complex environment by focusing on making decision for fixed channels. RL teaches an agent how to take optimal decisions in a complex environment by observing the agent from previous history (state-action reward pair) and agent will take the best action. The problem solving in complex environments has become easy with the development of deep neural network. In MDP, best action is selected by agent depending upon the current state. Mathematically, MDP is represented as

$$p[s_{t+1}/s_{t1}] = p[s_{t+1}/s_{t1} \cdots s_t] \tag{2}$$

where s_{t+1} represents the state at time step $t+1$. Transition probability from state s_t to s_{t+1} is given by $p[s_{t+1}/s_t]$. The above equation states that the previous state [s_t] has all the previous information from state [s_1—s_{t-1}] for making next decision. The present state consists of a set of state 's', action 'a', and reward 'r' which are the three fundamental elements in MDP. At time step t, the agent observes state 's_t ' and takes action 'a_t ', receives reward 'r_t ' and goes for the next state s_{t+1} [42]. In the proposed approach, LEO satellite acts as an agent. Maximizing the utility function, maximizes $\sum_{t=1}^{T} r_t$ where utility function is the user's satisfaction rate with the service provided by the agent.

3.2.1 State Space

In the DRL-LSTM approach, footprint of the satellite is divided into $M * N$ grid form. Each grid determines the user request at each time step. State space includes information about the present user requests and new user requests. We consider images *Y, Y+1* with $M * N$ form, where M and N represent width and height of the image respectively, which depends upon the coverage area of the satellite. Image Y refers to initial user requests and image $Y + 1$ represents new user request for the channels. Values in the grid are either 0 or 1. 0 means channel is free while 1 means channel is under use as shown in Fig. 6.

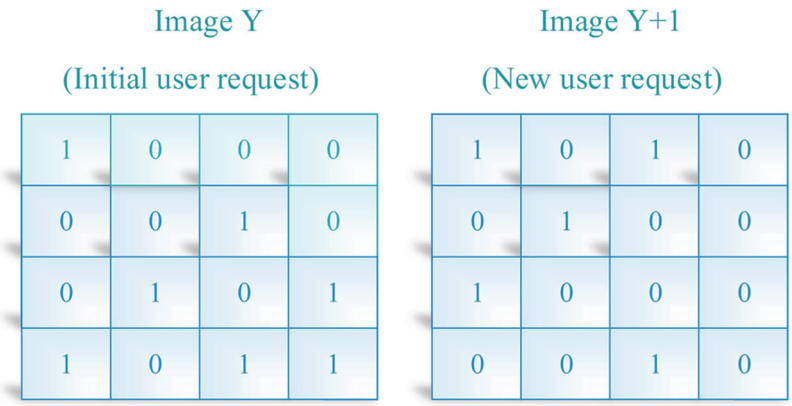

Figure 6: State representation

3.2.2 Action

Action (a_t) includes plotting of new user requests with the required channels. For each new user request, agent (satellite) allocates a channel from the available channels. There may be a possibility where no channel is allocated for serving the user (n = 0). In practical scenario, the users at each time step are very large. Hence, we require large action space, and therefore, learning becomes challenging. For reducing the action space, the action is divided into a set of mini-actions $A = [a_1, a_{2,}—a_i]$.

3.3.3 Reward

Reward (r_t) is the feedback that the agent receives after taking an action. That is, after each time step and after each reward, the satellite allocates a channel. Some important symbols and representation are shown in Tab. 1.

Table 1: Symbols and representations

Symbols	Representation
N_b	Total number of beams
N_c	Total number of channels
B_{tot}	Total bandwidth
P_{tot}	Total transmission power of satellite
P_b	Power allocated to each beam
P_c	Maximum transmission power of each channel
B_c	Bandwidth of each channel
T_b	Time duration of the satellite that covers each block
p_o	Power allocation of one episode
p_b	Power consumption in each beam
r_{blo}	Blocking rate
A	Action

4 Methodology

The proposed DRL-LSTM model is shown in Fig. 7. It takes categorically designed training samples to completely learn and allocate the channels efficiently. At each episode the agent gets the observation from the environment which includes the present state of the channel allocation and of the users along with the previous training samples. After the observation, the agent takes resource allocation through the LSTM neural network.

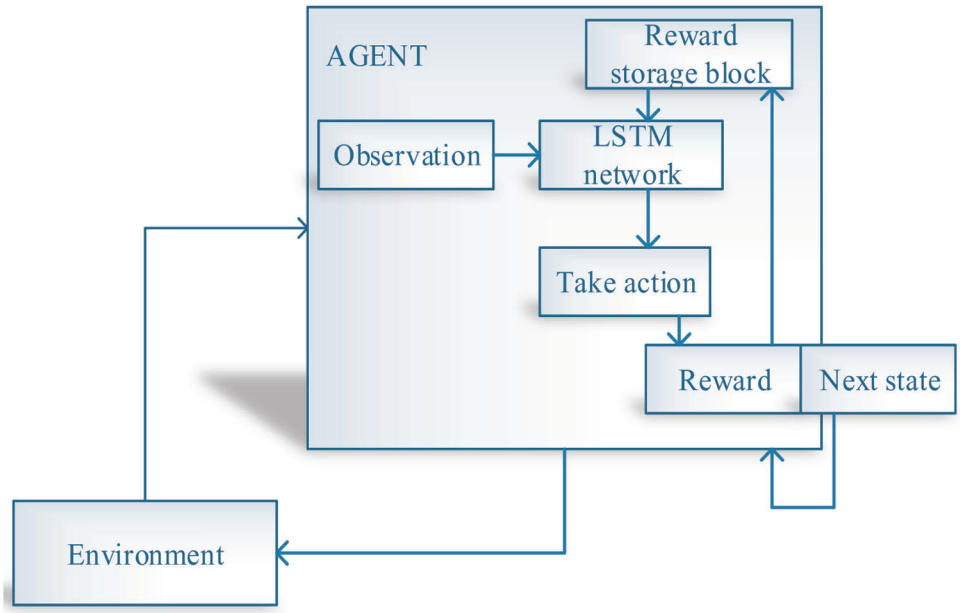

Figure 7: Proposed DRL-LSTM model

After resource allocation, the agent receives the reward which is stored in the reward storage box as experience replay for predicting the next state. Algorithm for the resource allocation is shown in Tab. 2.

Table 2: Algorithm for resource allocation in steps

S.No.	Resource allocation mechanism
1	Inputs: Output estimated matrix is given to agent.
2	Initialize the parameters: BW, no of channels, power
3	For episode e = 1: N Update the parameters based on the input
4	Observe the input and form the state s_t. Calculate the action set $A(s_t)$
5	**If** action A (s_t) is successful **do** Update the parameter from previous to present value Set immediate reward r_t
6	**Else** Reformulate the state $s_{t=}s_{t+1}$
7	Store the experience tuples.
8	With probability € select random action a_t€ A (s_t) Execute Channel Allocation action
9	Update the parameters bandwidth, power, no of channels
10	Set the reward

4.1 LSTM Concept

In order to understand the LSTM concept, one must be aware of the RNN concept. Both operate in a similar fashion with the exception that LSTM operates within the LSTM cell. Before going into the operation of the LSTM cell, let's take an example. Whenever we read the reviews of any online product which says, 'it is amazing', 'it works super', etc., then our brain concentrates on the words like 'amazing', 'super' and it doesn't care about 'it' and 'is' in the sentence. This means that our brain stores the relevant data and forgets the irrelevant data. This is what LSTM actually does. It stores the relevant information and forgets the irrelevant information and makes future predictions. In case of channel allocation, the LSTM cell stores the data of the user requests which are higher than the threshold value, skips the remaining information and updates the cell state with new values. The LSTM architecture is shown in Fig. 8.

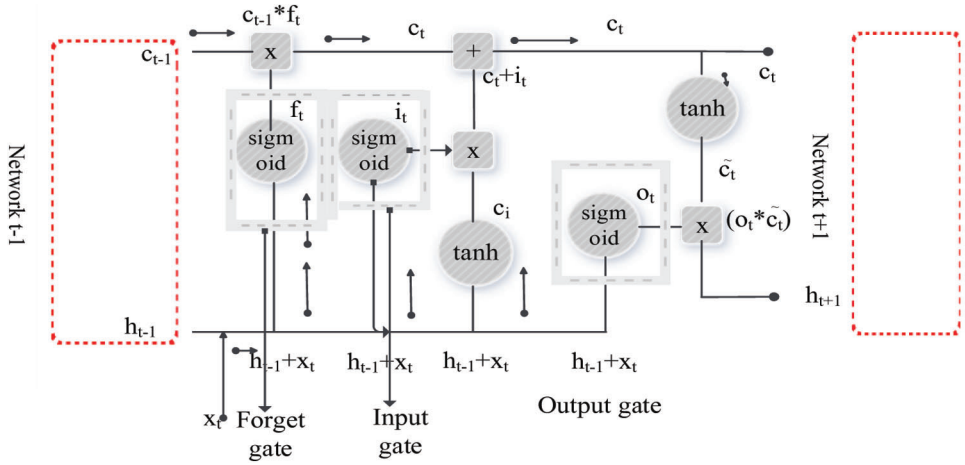

Figure 8: LSTM architecture

LSTM has a cell state and three gates. Cell state behaves as the transmission media that transmits the information throughout the cell. During the transfer of information to the entire cell either the information is added to the cell or removed from the cell by the gates. Gates are neural networks which are responsible for storing the relevant data during the training phase. Each neural network is associated with some weight *w*. There are two activation functions in the LSTM cell, *sigmoid* activation and *tanh* activation. The *tanh* activation function flattens the values between −1 and 1 whereas the *sigmoid* activation function flattens the values between 0 and 1. The following are the different types of gates in the LSTM cell.

4.1.1 Cell State

The cell state is used for calculation. Firstly, the previous cell state (c_{t-1}) information undergoes point wise multiplication with the output of the forget gate (f_t). It drops some of the values that are multiplied by zero (there is no resource allocation) and stores the values multiplied by 1(allocated resource) and updates the cell state from c_{t-1} to c_t. Next, it undergoes addition operation with the output of input gate and candidate value (c_i) and updates the new cell state with new values that the neural network finds relevant.

4.1.2 Forget Gate

In the forget gate, previous hidden state information (h_{t-1}) and current input (x_t) are allowed to pass through the *sigmoid* function. The output of the *sigmoid* function ranges between 0 and 1. Value that is closer to 0 indicates that the value should be forgotten and the value closer to 1 is to be stored. Hence,

the forget gate decides which information is to be forgotten from the cell state.

$$\text{Forget gate} \;=\; \text{Sigmoid}\left\{\sum_{t=1}^{N} w_f h_{t-1} + x_t + b_f\right\}$$

4.1.3 Input Gate

Input gate is used for updating the cell state. We first pass the input (x_t) and hidden cell state (h_{t-1}) value to the *sigmoid* function. The output of *sigmoid* function (i_t) will decide which values should be stored in the cell state by converting the values between 0 and 1. The input and hidden state (h_t+x_t) value is also passed to the *tanh* function to flatten the values between −1 and 1. Next, the output of *tanh* function (c_i) is multiplied with the *sigmoid* output. *Sigmoid* function decides what information is to be stored from the *tanh* output.

$$\text{Input gate} \;=\; \text{Sigmoid}\left\{\sum_{t=1}^{N} w_i h_{t-1} + x_t w_i + b_i\right\}$$

4.1.4 Output Gate

Output gate (o_t) predicts what would be the next hidden state (h_{t+1}) value. This hidden state contains the previous input values. Firstly, we pass the previous hidden information (h_t) and present input (x_t) to the *sigmoid* function and on the other hand the new cell state (c_t) values are passed to the *tanh* function. Now the output of the *tanh* function and the *sigmoid* function are multiplied and the output gives the next hidden state. The new cell state and new hidden state are forwarded to the next step for further processing.

$$\text{Output gate} \;=\; \text{Sigmoid}\left\{\sum_{t=1}^{N} w_o h_{t-1} + w_o x_t + b_o\right\}$$

4.2 Working

In DRL-LSTM, input is fed to the forget gate along with the previous hidden state (which is the normalized value). It is allowed to pass through a *sigmoid* function whose output lies between 0 and 1. This means the values which are closer to zero are multiplied by zero and the values closer to or greater than 1 are multiplied by 1. Output of *sigmoid* gives the forget gate output which undergoes point wise multiplication with the cell state and forgets the values which are multiplied by zero. Now the cell state is updated. This updated cell state contains the values ranging from 0 to 1(0 means no channel allocation and 1 means channel allocation). The same hidden state information and present input is fed to the input layer. This input layer undergoes *sigmoid* operation and output lies between 0 and 1. On the other hand, the same hidden state information and present input is fed to the *tanh* function which flattens the values between −1 and 1. This output of *tanh* function undergoes point wise multiplication with the *sigmoid* output and decides which values of *tanh* output are to be stored in the cell state (important information). Output from the multiplier undergoes point wise addition with previous cell state and updates the cell state information. This updated cell state values are allowed to pass through the *tanh* function. On the other hand, the previous hidden state information along with present input is given to the *sigmoid*. This *sigmoid* output undergoes point wise multiplication with output of *tanh* function and the output of this multiplier is the next hidden (predicted) state and updated cell state is given to the next time step. Algorithm of LSTM working is shown in Tab. 3.

Table 3: Algorithm for LSTM Working

S.No.	LSTM working algorithm
1	Input: Previous samples and present input is given to LSTM
2	Randomly initialize the weights (W's) and bias (b's) values
3	The forget gate: f_t = Sigmoid $\{\sum_{t=1}^{N} w_f h_{t-1} + x_t + b_f\}$
4	The input gate: I_t =sigmoid $\{\sum_{t=1}^{N} w_i h_{t-1}\, x_t w_i + b_i\}$
5	The candidate value: $\widetilde{ct}$=Tanh $\{\sum_{t=1}^{N} w_c h_{t-1} + w_c x_t + b_c\}$
6	Cell state is updated c_{t-1} to c_t
7	Cell state $c_t = \{c_{t-1} * f_t\} + \{i_t * \widetilde{ct}\}$
8	LSTM is updated by Output gate: O_t =Sigmoid $\{\sum_{t=1}^{N} w_o h_{t-1} + w_o x_t + b_o\}$
9	Output: Estimated next cell state c_{t+1}

5 Evaluation Results and Discussion

Extensive evaluation is carried out to evaluate the performance of the proposed DRL-LSTM scheme. Simulations are carried out using MATLAB version R2015a, which is modelled to test DRL-LSTM and other compared schemes under different ground situations while putting number of beams, total power, number of channels and some other parameters of the satellite constant. The simulation parameters used in the proposed work are shown in Tab. 4.

Table 4: Simulation parameters

Parameters	Quantity
Total power (dBW)	21
Total beams	25
Overall power of the satellite	25
Total channels	16
Maximum beam transmission power	20

5.1 Bandwidth Distribution

Fig. 9 depicts the distribution of bandwidth depending on the demand using Gaussian distribution. From Fig. 4, representing the node requests on the ground we consider that the square block moves with respect to the time slot from one column to the next column. Assume a column in Fig. 4 (4, 2, 6, 1, 8 …). Whenever the time slot of this column is finished, it moves to the next column, leaving the first column. We compare our method with various existing similar schemes like Greedy Algorithm (greedy 1, greedy 2), random power allocation (RPA) and Deep CA proposed in [41] .

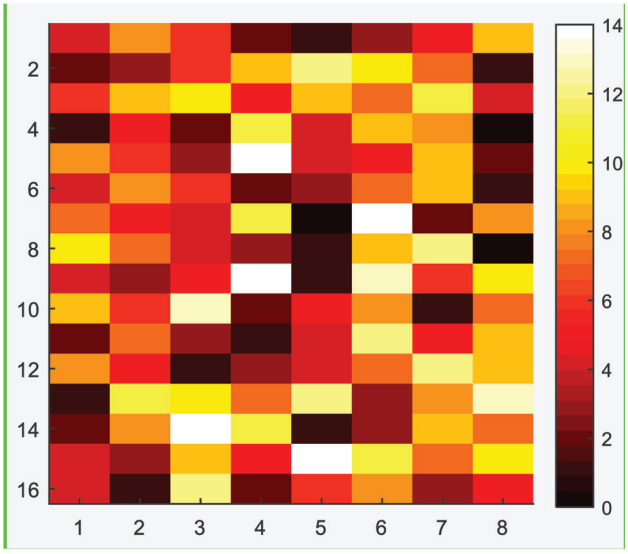

Figure 9: Distribution of bandwidth based on demand

5.2 Reward Function

Fig. 10 defines the reward function curve. We consider the user dataset to train our LSTM neural network. This dataset consists of the traffic of the user request for channel at each episode. The ground area is divided into a square grid of *M***N* which is further divided into mini squares of equal size. The dataset contains the user traffic information of each mini square for different episodes. The dataset consists of 1000 samples. Out of these we consider 800 samples for training and the rest for testing. We consider two cases of training the neural network (i) training the neural network till 350 episodes as shown in Fig. 10a and training the neural network till 550 episodes as shown in Fig. 10b. From Figs. 10a and 10b we can observe that there is improvement in the reward in Fig. 10b compared to 10(a). This is due to the reason that the neural network is given input for larger episodes so that it can analyze much better than the one with 350 episodes. This means larger the data, better the output. The allocation is approximately equal to the demand as we can see the increment in the reward. Due to the traffic variance the agent goes through some cases where the proposed model shows some fluctuation in the reward. However, this does not affect the performance of the model.

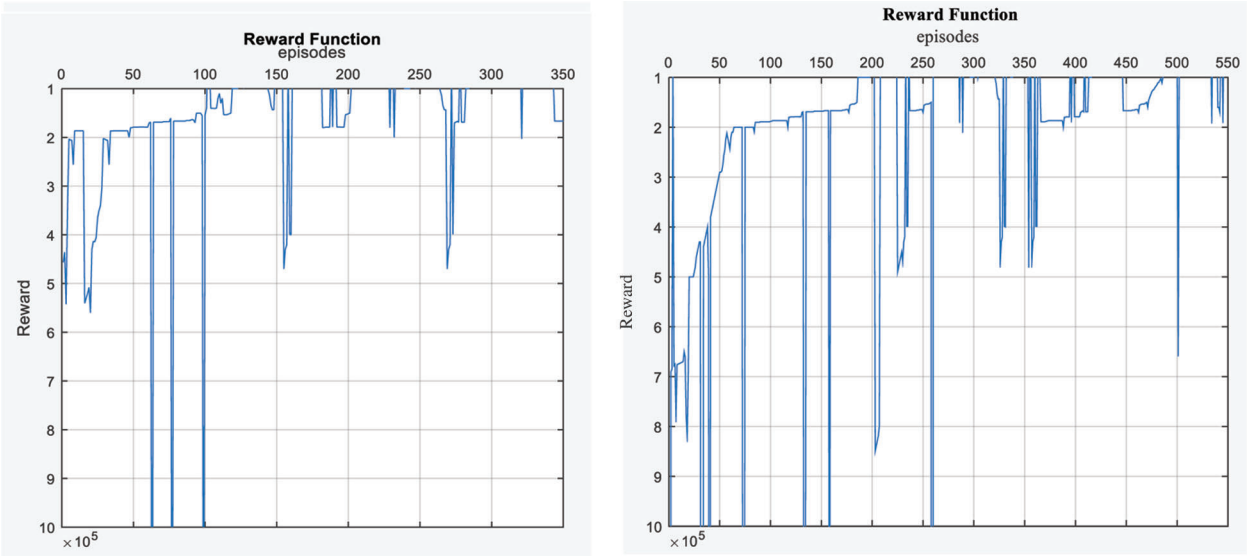

Figure 10: (a) Reward function plot for 350 episodes (b) Reward function plot for 550 episodes

5.3 Power Consumption

One of the objectives of this article is to enhance the power efficiency of the satellite and to provide quality of service (QoS) to users by reducing the blocking rate (r_{blo}). QoS is defined as overall performance of a network by providing specific requirements to the users with certain quality. QoS depends on the blocking rate probability. Blocking rate is defined as ratio of the number of rejected requests by the satellite to the overall requests from the ground due to the limited availability of the resources. To guarantee QoS to users, threshold value (br_{max}) should always be greater than (r_{blo}).

$$br_{max} > r_{blo} \tag{3}$$

This threshold value is defined as per the QoS requirements of the users.

While in training phase one episode comprises 1000 blocks. The power consumption P_c of satellite in each beam is defined as P_b at time step P_t. The total power consumption for one episode is defined by $\sum_{t=1}^{T} P_t$. In this case, power is divided into M*N matrix which is denoted by P_{Bm}, P_{Bn}. Here the power allocated to each beam is greater than the power of the matrix. P_o denotes power allocation of one episode.

$$P_c = \frac{\sum_{t=1}^{T} P_t - P_o}{P_o} \tag{4}$$

Fig. 11 shows the result of the power consumed by all the five schemes and efficient working of DRL-LSTM model over the other schemes. This efficiency is achieved by training the LSTM with previous user requests information in each block along with present input so that it is aware of the user requests in each block and allocates channels based on requirement to reduce the consumption of power. In Fig. 11, for average demand density, we observe that the power consumed by greedy 2 (GA_2) algorithm is very low but the blocking rate is very high which can be seen in Fig. 12. This means that it provides energy efficient service but the satisfaction rate is low which is high in our approach. Satisfaction rate is expressed as the ratio of users that have been allocated channels to the sum of overall users. Here the satisfaction rate is mirrored by blocking rate. Lesser is the blocking rate, more is the satisfaction rate.

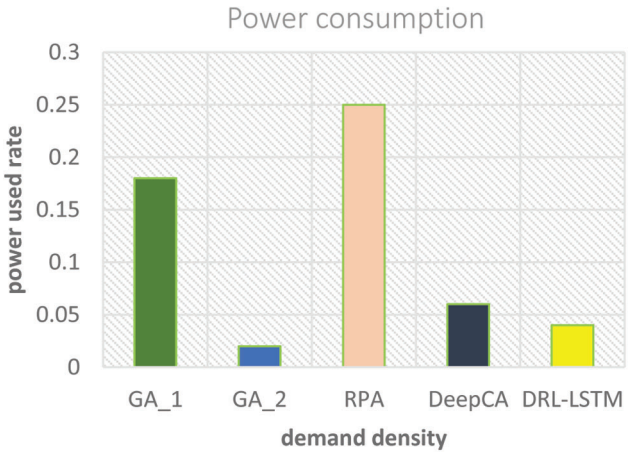

Figure 11: Power consumption comparison of our proposed method with different demand distribution

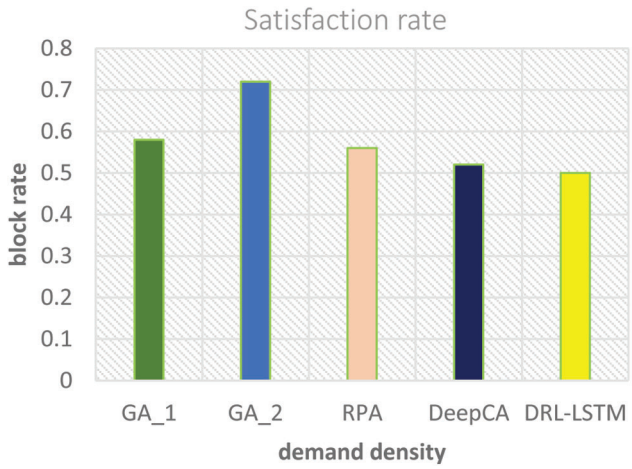

Figure 12: Blocking rate comparison of DRL-LSTM with other schemes

5.4 Satisfaction Rate

In the proposed DRL-LSTM approach, the rate of satisfaction is higher compared to all the other four schemes. The blocking rate in our approach is slightly lesser than Deep CA. The results of Figs. 11 and 12 reveal that DRL-LSTM provides better result as compared to other schemes by reducing power consumption with high satisfaction rate. This result is achieved by training the LSTM network with larger previous information over longer periods instead of short time so that it is capable of allocating the channels efficiently with minimizing power consumption. Due to this, the agent performs best actions and achieves best rewards. Tab. 5 summarizes the compared schemes based on power consumption and blocking rate.

Table 5: Comparison of different schemes with proposed approach

Parameters	Greedy-1	Greedy-2	RPA	DeepCA	DRL-LSTM
Power Consumption	0.18	0.02	0.25	0.06	0.04
Blocking Rate	0.58	0.72	0.56	0.52	0.5

6 Conclusion

In this article, we proposed a DRL-LSTM based novel approach for dynamic channel allocation in SIoT based on the demand. We executed the entire footprint of the satellite as a square matrix for serving the resources of the satellite to the ground users by considering the channel allocation problem as MDP. We used LSTM neural network for storing the previous user requests from the ground and predicting the channel requirement. Our approach is evaluated by extensive simulation and is found to perform better in providing high satisfaction rate and low power consumption compared to other existing classical channel allocation schemes. With the proposed DRL-LSTM model, the power consumption is reduced by 0.02% with respect to DeepCA and blocking rate is reduced by 0.02% than DeepCA; by 0.06% than RPA; by 0.22% than Greedy 2; and by 0.08% than Greedy 1 algorithms.

Acknowledgement: This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, Saudi Arabia. The authors, therefore, gratefully acknowledge the DSR technical and financial support.

Funding Statement: The authors received no specific funding for this study.

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

Reference

[1] C. Wang, H. Wang and W. Wang, "A two-hops state-aware routing strategy based on deep reinforcement learning for LEO satellite networks," *Electronics*, vol. 8, no. 9, pp. 920, 2019.

[2] J. Wei, J. Han and S. Cao, "Satellite IoT edge intelligent computing: A research on architecture," *Electronics*, vol. 8, no. 11, pp. 1247, 2019.

[3] S. Cho, I. F. Akyildiz, M. D. Bender and H. Uzunalioglu "A new connection admission control for spotbeam handover in LEO satellite networks," *Wireless Networks*, vol. 8, pp. 403–415, 2002.

[4] M. I. Nazeer, G. A. Mallah and R. A. Memon, "A hybrid scheme for secure wireless communications in IoT," *Intelligent Automation & Soft Computing*, vol. 29, no. 2, pp. 633–648, 2021.

[5] Future IOT Editors, "Satellites to enable 24 million IoT connections globally by 2024," July 15, 2019.

[6] Z. Qu, G. Zhang, H. Cao and J. Xie, "LEO satellite constellation for internet of things," *IEEE Access*, vol. 5, pp. 18391–18401, 2017.

[7] I. Leyva-Mayorga, B. Soret, M. Roper, D. Wubben, B. Matthiesen *et al.*, "LEO Small-satellite constellations for 5G and beyond-5G communications," *IEEE Access*, vol. 8, pp. 184955–184964, 2020.

[8] P. Wang, J. Zhang, X. Zhang, Z. Yan, B. G. Evans *et al.*, "Convergence of satellite and terrestrial networks: A comprehensive survey," *IEEE Acess*, vol. 8, pp. 5550–5588, 2019.

[9] F. Zheng, Z. Pi, Z. Zhou and K. Wang, "LEO satellite channel allocation scheme based on reinforcement learning," *Mobile Information Systems*, vol. 22, pp. 10, 2020.

[10] C. Zhou, W. Wu, H. He, P. Yang, F. Lyu *et al.*, "Deep reinforcement learning for delay-oriented IoT task scheduling in SAGIN," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 911–925, 2021.

[11] C. Jin, X. He and X. Ding, "Traffic analysis of LEO satellite internet of things," in *2019 15th Int. Wireless Communications & Mobile Computing Conf. (IWCMC)*, IEEE, Tangier, Morocco, pp. 67–71, 2019.

[12] M. A. A. Madni, S. Iranmanesh and R. Raad, "DTN and non-dTN routing protocols for inter-cubesat communications: A comprehensive survey," *Electronics*, vol. 9, no. 3, pp. 482, 2020.

[13] J. A. Fraire, S. Céspedes and N. Accettura, "Direct-to-satellite IoT—a survey of the state of the art and future research perspectives: Backhauling the IoT through LEO satellites," in *International Conference on Ad-Hoc Networks and Wireless*, Luxembourg, Luxembourg, pp. 241–258, 2019.

[14] L. Polak and J. Milos, "Performance analysis of loRa in the 2.4 GHz ISM band: Coexistence issues with Wi-fi," *Telecommunication Systems*, vol. 40, no. 3, pp. 299–309, 2020.

[15] C. Han, A. Liu, L. Huo, H. Wang and X. Liang "A prediction-based resource matching scheme for rentable leo satellite communication network," *IEEE Communications Letters*, vol. 24, no. 2, pp. 414–417, 2019.

[16] B. Beng, C. Jiang, H. Yao, S. Guo and S. Zhao, "The next generation heterogeneous satellite communication networks: Integration of resource management and deep reinforcement learning," *IEEE Wireless Communications*, vol. 27, no. 2, pp. 105–111, 2020.

[17] J. Wang, L. Zhao, J. Liu and N. Kato, "Smart resource allocation for mobile edge computing: A deep reinforcement learning approach," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 3, pp. 1529–1541, 2019.

[18] F. Rasheed, K. L. A. Yau and Y. C. Low, "Deep reinforcement learning for traffic signal control under disturbances: A case study on sunway city, Malaysia," *Future Generation Computer Systems*, vol. 109, pp. 431–445, 2020.

[19] V. Mnih, K. Kavukcuoglu, D. Silver, A. Rusu, J. Veness *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[20] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez *et al.*, "Continuous control with deep reinforcement learning," in *6th International Conference on Learning Representations (ICLR 2016)*, San Juan, Puerto Rico, pp. 1–14, 2015.

[21] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap *et al.*, "A synchronous methods for deep reinforcement learning," in *The 33rd International Conference on Machine Learning*, New York City, NY, USA, pp. 1928–1937, 2016.

[22] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou *et al.*, "Playing atari with deep reinforcement learning," in *NIPS Deep Learning Workshop 2013*, Stateline, NV, USA, 2013, arXiv:1312.5602.

[23] R. Trub and L. Thiele, "Increasing throughput and efficiency of LoRaWAN class a," in *UBICOMM 2018. The Twelfth Int. Conf. on Mobile Ubiquitous Computing, Systems, Services and Technologies*, Athens, Greece, pp. 54–64, 2018.

[24] A. Augustin, J. Yi, T. Clausen and W. M. Townsley, "A study of loRa: Long range & low power networks for the internet of things," *Sensors*, vol. 16, no. 9, pp. 1466, 2016.

[25] Y. Wang, J. Yang, X. Guo and Z. Qu, "Satellite edge computing for the internet of things in aerospace," *Sensors*, vol. 19, no. 20, pp. 4375, 2019.

[26] S. K. Routray and H. M. Hussein, "Satellite based IoT networks for emerging applications," arXiv:1904 00520, 2019.

[27] S. Cho, "Adaptive dynamic channel allocation scheme for spotbeam handover in LEO satellite networks," in *Vehicular Technology Conf. Fall 2000, IEEE VTS Fall VTC2000. 52nd Vehicular Technology Conf. (Cat. No. 00CH37152)*, vol. 4, pp. 1925–1929, 2000.

[28] J. Jiao, Y. Sun, S. Wu, Y. Wang and Q. Zhang, "Network utility maximization resource allocation for NOMA in satellite-based internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3230–3242, 2020.

[29] X. Zhang, D. Guo, K. An, G. Zheng, S. Chatzinotas *et al.*, "Auction-based multichannel cooperative spectrum sharing in hybrid satellite-terrestrial IoT networks," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 7009–7023, 2020.

[30] J. Jiao, Y. Sun, S. Wu, Y. Wang and Q. Zhang. "Network utility maximization resource allocation for NOMA in satellite-based internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3230–3242, 2020.

[31] A. M. Lewis and S. V. Pizzi, "Quality of service for tactical data links: TDMA with dynamic scheduling," in *MILCOM 2005-2005 IEEE Military Communications Conf.*, vol. 4, pp. 2350–2359, 2005.

[32] S. Ilcev, "Implementation of multiple access techniques applicable for maritime satellite communications," *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 7, no. 4, pp. 529–540, 2013.

[33] I. Lysogor, L. Voskov, "Study of data transfer in a heterogeneous loRa-satellite network for the internet of remote things," *Sensors*, vol. 19, no. 15, pp. 3384, 2019.

[34] S. Liu, X. Hu and W. Wang, "Deep reinforcement learning based dynamic channel allocation algorithm in multibeam satellite systems," *IEEE Access*, vol. 6, pp. 15733–15742, 2018.

[35] A. J. Onumanyi, A. M. Abu-mahouz and G. P. Hancke, "Low power wide area network, cognitive radio and the internet of things: Potentials for integration," *Sensors*, vol. 20, no. 23, pp. 6837, 2020.

[36] R. M. Sandoval, A. J. Garcia-Sanchez and J. Garcia-Haro, "Optimizing and updating lora communication parameters: A machine learning approach," *IEEE Transactions on Network and Service Management*, vol. 16, no. 3, pp. 884895, 2019.

[37] M. Jia, L. Jiang, Q. Guo, X. Jing, X. Gu *et al.*, "A novel hybrid access protocol based on traffic priority in space-based network," *IEEE Access*, vol. 6, pp. 24767–24776, 2018.

[38] R. Prasad and Shivashankar, "Improvement of battery lifetime of mobility devices using efficient routing algorithm," *Asian Journal of Engineering Technology and Applications*, vol. 1, no. 1, pp. 13–20, 2017.

[39] L. Xu, G. M. P. O'Hare and R. Collier, "A smart and balanced energy-efficient multihop clustering algorithm (smart-beem) for mimo iot systems in future networks," *Sensors*, vol. 17, no. 7, pp. 1574, 2017.

[40] F. Chiti, R. Fantacci and L. Pierucci, "Energy efficient communications for reliable IoT multicast 5gsatellite services," *Future Internet*, vol. 11, no. 8, pp. 164, 2019.

[41] B. Zhao, J. Liu, Z. We and I. You, "A deep reinforcement learning based approach for energy-efficient channel allocation in satellite internet of things," *IEEE Access*, vol. 8, pp. 62197–62206, 2020.

[42] D. Wang, H. Qin, B. Song, K. Xu, X. Du *et al.*, "Joint resource allocation and power control for D2D communication with deep reinforcement learning in MCC," *Physical Communication*, vol. 45, pp. 101262, 2021.