Efficient Load Balancing for Future Dense Networks using Radio over Fiber Infrastructure and applying Different Learning Rates

Mahfida Amjad Dipa
Wireless and Photonics Networks Center, Faculty of Engineering, Universiti Putra Malaysia, Serdang, Malaysia
amjad.mahfida@student.upm.edu.my (corresponding author)

Syamsuri Yaakob
Wireless and Photonics Networks Center, Faculty of Engineering, Universiti Putra Malaysia, Serdang, Malaysia
syamsuri@upm.edu.my

Fadlee Rasid
Wireless and Photonics Networks Center, Faculty of Engineering, Universiti Putra Malaysia, Serdang, Malaysia
fadlee@upm.edu.my

Faisul Ahmad
Wireless and Photonics Networks Center, Faculty of Engineering, Universiti Putra Malaysia, Serdang, Malaysia
faisul@upm.edu.my

Azwan Mahmud
Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia
azwan.mahmud@mmu.edu.my

Received: 3 January 2025 | Revised: 3 February 2025 | Accepted: 11 February 2025

 DOI: https://doi.org/10.48084/etasr.10122

ABSTRACT

Reinforcement Learning (RL) can lead to effective Load-Balancing (LB) mechanisms, as traditional methods cannot always provide an optimal solution in cellular networks. This study proposes an RL-based LB scheme for a dense network that uses radio over fiber infrastructure. The proposed technique is based on LB constraints in the action space that maintain zero violation during the learning process. In this technique, a Deep Q-Network agent was chosen to search for an optimal policy to maximize the expected cumulative long-term reward to satisfy the constraints. This study uses the number of user entities per base station in the dense network as constraints to maintain average throughput based on the Signal-to-Noise Ratio (SNR) generated from the radio frequency signals of the network. The proposed method outperformed at an SNR of 38 dB with a throughput of 32 Mbps for a 20 MHz channel bandwidth for macro- and microcells in the dense network. Furthermore, this study examined the effect of different learning rates as hyperparameters in the system. The proposed approach shows that when the agent was trained with a learning rate of 1e-3, the network performed well by obtaining a higher CDF compared to a learning rate of 1e-5. In addition, the system achieved higher rewards for a learning rate of 1e-3 with or without LB constraints, confirming the efficiency of the proposed scheme. The simulation results showed that CDF was 4% higher when using constraints compared to without constraints.

Keywords-RoF; dense network; radio over fiber; load balancing; reinforcement learning; learning rate

I. INTRODUCTION

Cellular networks can be uneven when the cell has an unconventional traffic distribution. An overloaded cell cannot satisfy user requirements and can cause problems with throughput, delay, blocking, and drop [1]. In addition, there is a waste of network resources for the underloaded cell. A relatively small network coverage can reduce the system's performance due to the frequent movement of the users. Considering the issue of small cell coverage with a high traffic load, a Load Balancing (LB) scheme is very crucial to address the problems of traffic load distribution in dense cellular networks [2].

Radio over Fiber (RoF) technology has gained a lot of interest in handling massive data traffic [3, 4]. A RoF network is a hybrid optical wireless network that allows for extremely high bandwidth, and due to the characteristics of the wireless network, user mobility is not affected [5]. These networks can have a hundred times more traffic handling capacity than 4G networks, gigabit service capability per user, low latency, and high spectral efficiency [6-8].

In conventional LB techniques, researchers are using different rule-based techniques, such as the swap-based LB approach between access points [9], multi-depth offloading algorithm considering the radio resource allocation [10], and metaheuristic algorithm, in radio access networks [11]. In addition, new techniques such as Reinforcement Learning (RL) have shown effectiveness in solving LB issues in communication [12]. RL attempts to learn control policies by interacting with the environment. However, RL-based techniques have inherent challenges. RL requires frequent interactions with the environment to learn a satisfactory policy and a reward function to achieve the desired performance [12]. Based on a distributed multi-agent Deep Q-network (DQN), the LB mechanism in [13] focused on user association for dense networks. It used a matching game-based policy for LB, where each base station maintained a preference list to make association decisions using the sum data rate and a learning rate for the system of 0.001. Another RL-based LB scheme was presented in [14] for a heterogeneous LiFi WiFi network, using three different RL reward functions to maximize average network throughput and user satisfaction in terms of user data rate, with a learning rate of 0.01.

Although different LB techniques were reported in [11, 15-18] for different networks, to the best of our knowledge, LB for RoF infrastructure based on RL utilizing DQN and focusing on different learning rates has not been reported yet. Therefore, this study proposes an LB technique using an RL algorithm for the RoF infrastructure that adopts a DQN agent. The objectives of this study can be summarized as follows:

- Proposes an RL-LB technique for dense networks based on RoF infrastructure.
- Utilizes a DQN agent for the RL-LB scheme.
- Applies two different learning rates to evaluate the performance of the RL-LB scheme.

II. ROF NETWORK ARCHITECTURE

A fiber wireless-based dense RoF system was designed by adopting the network parameters of 3GPP TR 38.913 [19]. For simulating the dense cellular network, two cells, macro and microcell, were used, operating at 4 GHz and 30 GHz RF signals. The RoF dense network was simulated using Optisystem v21, and Figure 1 shows its schematic diagram. The control station consists of a Laser Diode (LD) with a wavelength of 1550 nm to generate continuous optical light waves, producing an output power of -10 dBm. The signal was transmitted to a Li-Nb-MZM, driven by a 2 GHz and a 15 GHz RF signal. It was then launched into a 10 km Single Mode Fiber (SMF) with a loss of α = 0.21 dB/km, and chromatic dispersion D = 16 ps/nm/km. At the base station, the optical signal was detected by a high-speed PD with a 3 dB bandwidth of 4 GHz and 30 GHz and a responsivity of 0.6 A/W. The PD output was fed into a spectrum analyzer to be monitored and measured. To achieve transmission at the minimum bias point, VDC1 was set to 4 V, VDC2 was set to 0 V, and the LO modulation index was 0.7. At the receiving end, the optical receiver consisted of a PIN Photodetector (PD) to convert the optical field to current, followed by a Transimpedance Amplifier (TIA) for amplification, and then demodulated by Amplitude (AM) demodulators for obtaining the electrical signals. The generated photocurrent was analyzed by the RF spectrum analyzer and the BER analyzer showing the BER performance. Table I shows the design parameters of the dense network environment.

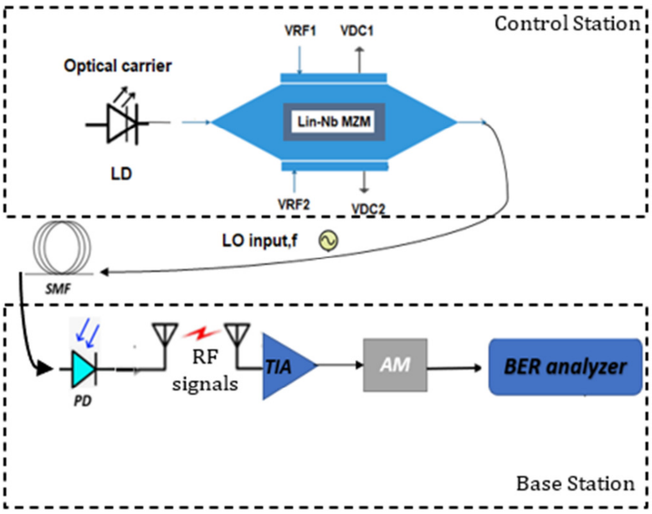

Fig. 1. Schematic diagram of the RoF network..

TABLE I. DESIGN PARAMETERS OF DENSE NETWORK

Parameters	Values
RF signals	4 GHz (Macrocell) 30 GHz (Microcell)
Number of UE	10 UE per BS
UE distribution	Uniform random distribution
Threshold value of SNR	10 dB
Channel bandwidth, BW	20 MHz

The RoF system can be evaluated utilizing several performance-analyzing characteristics and factors, such as Signal to Noise Ratio (SNR), Bit Error Rate (BER), and Quality (Q) factors to find out whether or not the system is operating appropriately or not [20]. This study evaluated the performance of the RoF dense network in terms of SNR, which was calculated by comparing the received signal power with the noise power of the RF signal. The following equation was adopted to calculate SNR [21, 22]:

$$\gamma = \Re - \mathbb{N} \quad (1)$$

where γ, $\Re$, and $\mathbb{N}$ represent SNR (dB), received signal power (dBm), and noise power (dBm), respectively.

III. PROPOSED RL-LB METHOD

LB aims to balance the load across available network resources to improve network performance. Most existing LB techniques are designed and tuned manually, where near-optimality is difficult to achieve. Additionally, rule-based methods are difficult to adapt to rapidly changing traffic patterns in real-world environments. RL algorithms gain success in many application domains because of their adaptability to dynamic changes in network load patterns [12]. This section presents the LB method for the RoF dense network and the RL-LB technique. This scheme is based on a DQN agent. Figure 2 shows the block diagram of the RL-LB technique.

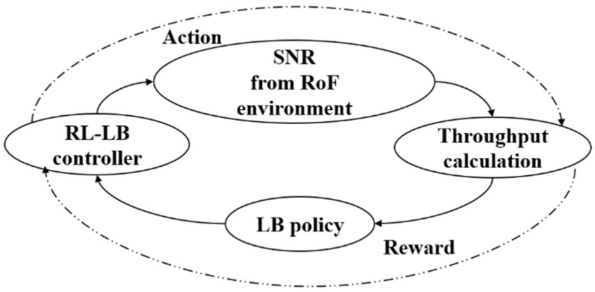

Fig. 2. A block diagram of the RL-LB scheme.

The goal is to evaluate the performance of the RL-LB controller in the RoF network environment. The main objective of the controller is to balance the load on the network. In this study, the number of User Entities (UE) under a Base Station (BS) was treated as load in the network. As DQN operates on a discrete action space, so in the RL-LB scheme, the action space has to be discretized. Two types of actions were modeled in the scheme: (i) the number of UE is within the range, and (ii) the number of UE is out of range, considering the acceptable SNR value between the UE and BS. The load constraint of the proposed LB technique is the number of UE per BS. By restricting the number of maximum UE at each BS, it can balance loads among all BSs in a dense wireless network [14].

Throughput was considered as a performance metric, measured with [14, 23-25]:

$$\Gamma = BW * log\ (1 + \gamma) \quad (2)$$

where Γ represents throughput, γ denotes the SNR derived from (1), and BW is the channel bandwidth (20 MHz). The agent, state, action, and reward of RL-LB are presented in the following section.

A. Agent

A single agent A, which is either a macrocell or microcell BS, learns the condition of the SNR of the UE from (1). Hence, the agent in RL-LB is UE per BS and denoted as Ψ, where:

$$\Psi \in \{1,\ 2, 3, \ldots, 10\};\quad [1 < \Psi < 10] \quad (3)$$

B. State

The state S represents the condition of a Ψ within the cell based on the γ. The set of S of a Ψ can be represented as

$S \in \{0, 1\}$.

If $\gamma > 1$ then $S = 0$, and if $\gamma < 40$ then $S = 1$ (4)

C. Action

The action At is based on the calculated Γ using (2). The set of At can be represented as

$At \in \{0, 1\}$.

If $\Gamma > 1$ then $At = 0$, and if $\Gamma < 1$ then $At = 1$ (5)

D. Reward

Reward R is defined in (6), inspired by [26], and controls the learning process to achieve the acceptable throughput (Γ). R has a maximum value of 1 when the Γ is exactly at the maximum and reduces to 0 when it is away from it.

$$R = [1 - |10 * (\Gamma)|] \quad (6)$$

After each trial, A updates the mapping to maximize the reward. This process continues until A learns an optimal mapping to maintain a balanced network.

Algorithm 1 presents the working steps of the proposed RL-LB approach. An RL policy is a mapping of an environment observation to a probability distribution of the actions to be executed. A value function is a mapping from an environment observation to a policy value. The policy value is defined as its expected discounted cumulative long-term reward. Agents use parameterized policies and value functions, which are implemented by function approximators called actors and critics, respectively. During training, the actor learns the policy that selects the best action to take. It does so by tuning its parameters to assign a higher probability to actions that yield higher values. The critic learns the value function that estimates the value of the current policy. The DQN learning policy is a discrete-action stochastic policy, which returns stochastic actions given an input observation, according to a probability distribution.

```
Algorithm 1 RL-LB scheme
Initialize the agent A based on (3)
For each episode:
  Reset the environment based on (4)
  Get the initial observations (S0) from
  the environment
  Calculate the initial action At0 based
```

```
on (5)
Set the current At to the initial action
(At ← At0), and set the current
observation to the initial observation
(S ← S0).
Initialize the critic Q(S,A;ϕ), φ =
random parameter values
Initialize the actor π(S;θ), θ = random
parameter values
While the episode is not finished or
terminated, perform the following steps:
  For the current observation S, select
  action At = π(S;θ)
  Apply At to the environment and obtain
  the next observation S′ and the reward
  R using (6)
  Compute the next At′
  Update the current At with the next
  action (At ← At′) and update the current
  observation with the next observation
  (S ← S′)
Terminate the episode if the termination
conditions defined in the environment
are met. Otherwise, begin the next
episode.
```

The RL-LB scheme was simulated on Matlab R2023b, and Table II presents the simulation parameters. Each agent's RL structure is constructed with three hidden layers with the Rectified Linear Unit (ReLU) activation function, with 64, 32, and 32 hidden nodes, respectively. In each time step, the weights of RL-LB are updated using the Adam optimizer with a mini-batch of size 64, and the weights are updated at each time step for training steps of 1000. The learning rate α is the most important hyperparameter while configuring the RL network. Hence, it is important to know how to investigate the effects of α on model performance [27]. The learning rates of RL-LB were chosen based on [13, 28] as 1e-3 and 1e-5, respectively, with a discount factor γ = 0.99. However, the default learning rate for the ADAM optimizer is 1e-3 [29].

TABLE II. SIMULATION PARAMETERS OF RL-LB SCHEME

Parameters	Values
Rectified Linear Unit (ReLU)	3 layers
Activation function of 3 layers	(64, 32, 32)
Learning rate, α	1e-5, 1e-3
Optimizer	Adam
Mini batch size	256
Number of episodes	500
Time steps per episode	1000
Discount factor, γ	0.99
Gradient decay	0.9
Gradient threshold	1

IV. PERFORMANCE ANALYSIS

This study examined how the number of UE performs based on SNR maintaining throughput in the system while satisfying the LB constraint. The learning was carried out in episodes, where each episode contains multiple learning time steps. During each episode, the DQN is learning and updating every time step, and each UE is carried out once per episode.

Figure 3 shows the system performance in terms of throughput with respect to SNR (1) for two cells. The graph shows that the throughput of the cells gradually increases with increasing SNR. At 10 dB SNR, the throughput for two cells is within 20 Mbps. The proposed method outperforms at an SNR value of 38 dB with a throughput of 32 Mbps for 20 MHz channel bandwidth for macro and microcells in the dense network. This behavior suggests that throughput increases with the increase of SNR, achieving good system performance [30].

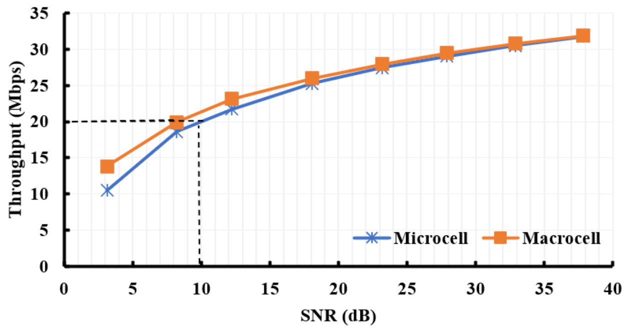

Fig. 3. Performance analysis: Throughput to SNR.

To validate the RL-LB scheme, the correlation between the Cumulative Distribution Function (CDF) and the number of episodes was examined. CDF is a statistical method to describe the distribution of random variables for a given set of parameters. CDF was defined by [31, 32]:

$$f\,(x,\mu,\sigma) \;=\; 1/\sqrt{(2\pi)}e^{-(x-\mu)2/2\sigma 2} \tag{10}$$

where μ is the mean of the distribution, σ^2 is the variance and x is the set of values. The CDF values range over the interval [0, 1] [33].

Furthermore, the performance of RL-LB was examined using two values for α as a hyperparameter of the system. A high value of α implies the willingness of the agent to learn from its environment and tends to yield higher rewards [34].

Figure 4 shows the correlation between the CDF and the number of episodes. The CDF of the user load becomes higher in each episode step. This situation is observed from CDF values of 0.3 to 1, for episodes 1 to 500, respectively. The values are gradually increasing in each episode and as a result, this verifies that the distribution of user load in the network is balanced [34]. It is also visible that the learning rate of 1e-3 performed better compared to 1e-5.

Figure 5 shows the performance analysis of the RL-LB scheme in terms of rewards (6). This figure shows the average rewards of the system for the two learning rates. The learning rate of 1e-3 obtained a higher reward compared to 1e-5. The two curves show how the model can learn the system well with a learning rate of 1e-3, and slower with a learning rate of 1e-5 [27]. With the designed model configuration, the results suggest a moderate learning rate of 1e-3 results in good model performance on the RL-LB scheme. The graph also shows the impact of LB constraints for the proposed RL-LB scheme according to (3) and (4). The numerical analysis shows that with or without LB constraints, again the learning rate 1e-3 achieves better performance than 1e-5.

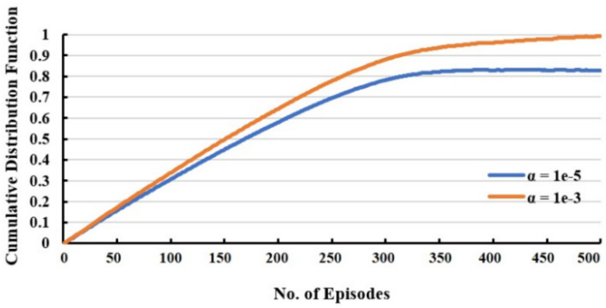

Fig. 4. Performance analysis of RL-LB: CDF vs number of Episodes.

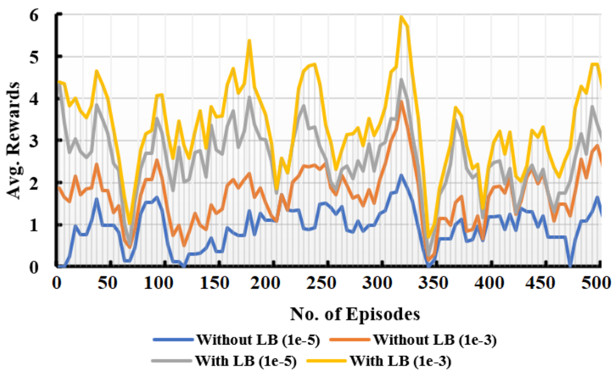

Fig. 5. Performance analysis of RL-LB scheme: Rewards vs number of Episodes.

Finally, Figure 6 shows the impact of the RL-LB scheme with or without LB constraints for a learning rate of 1e-3. The graph presents that the higher value of the CDF curve was around 0.86 for LB constraints and 0.86 without LB constraints. Therefore, it can be stated that using LB constraints, the proposed RL-LB scheme obtained a 4% higher CDF.

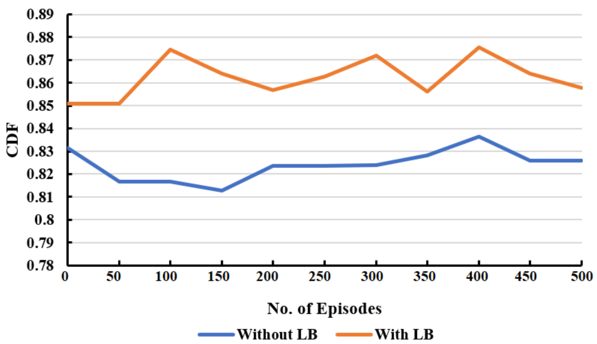

Fig. 6. Performance analysis of the proposed RL-LB scheme with and without using the LB constraints.

The proposed RL-LB scheme was compared with two recent approaches, as shown in Tables III and IV.

TABLE III. COMPARATIVE ANALYSIS

Method	Application	Design parameters	Performance metrics
Multi-Agent Deep Q-Learning [13]	Dense network	Sum data rate, signal-to-interference plus noise ratio (SINR)	Spectral efficiency, time steps
RL method with 3 reward functions [14]	LiFi/WiFi network	Signal-to-Noise Ratio (SNR)	Throughput, Complementary Cumulative Distribution Function (CCDF)
RL-LB (proposed)	RoF dense network	Signal-to-Noise Ratio (SNR)	Throughput, Cumulative Distribution Function (CDF), average rewards

TABLE IV. QUALITATIVE COMPARISON OF NETWORK DESIGN COMPLEXITY

Method	Computation complexity	Time complexity
Multi-Agent Deep Q-Learning [13]	Complex	High
RL method with 3 reward functions [14]	Medium	Medium
RL-LB (proposed)	Medium	Medium

V. CONCLUSION

This study presented an RL-based LB scheme for a dense RoF network. To improve efficiency, this study adopted deep learning using a DQN agent. The agent searches for an optimal policy that maximizes the expected cumulative long-term reward to satisfy the LB constraint. The numerical results show that the proposed RL-LB scheme provides comparable performance in terms of SNR and throughput. The proposed RoF network can obtain an acceptable SNR value, and as the SNR increases, throughput also increases, resulting in better system performance. The results showed that the network can achieve an SNR value of 10 dB, which is required for efficient signal transmission, and achieve a throughput of 20 Mbps. The proposed method achieved an SNR value of 38 dB with a throughput of 32 Mbps for 20 MHz channel bandwidth for macro and microcells in the dense network. Furthermore, the RL-LB approach shows that, when the agent is trained with a learning rate of 1e-3, the network performs well by obtaining a higher CDF compared to a learning rate of 1e-5. In addition, it exhibits similar behavior in terms of system rewards for two learning rate scenarios. Finally, this study suggests the learning rate as 1e-3 for the RL-LB scheme, as it provides good results. Using the learning rate of 1e-3, the proposed scheme was evaluated with and without LB constraints for CDF per number of episodes. The results show that the CDF value was 4% higher when using LB constraints compared to without LB constraints.

This work is limited to software simulation only and does not include any experimental analysis. In the future, the proposed method may offer an LB learning strategy in heterogeneous networks considering traffic patterns in real-world experiments. The technique of employing varying learning rates for various optimizers, such as stochastic gradient descent with momentum and root mean squared propagation, in addition to the adaptive movement estimation optimizer, will be further examined in the following phase. Future research could examine various optimizations to evaluate network scalability, including performance criteria such as block error rate and fairness index in the network.

VI. ACKNOWLEDGMENT

The authors acknowledge the sponsorship by the Organization for Women in Science for the Developing World (OWSD), Fund Reservation No. 3240318596, and the Fundamental Research Grant (FRGS), Ministry of Higher Education, Malaysia, No. FRGS/1/2023/TK07/UPM/02/11.

REFERENCES

[1] Z. He and J. Wang, "Non-cooperative differential game based Load Balancing Algorithm in Radio-over-Fibre system," *China Communications*, vol. 11, no. 14, pp. 79–85, 2014, https://doi.org/10.1109/CC.2014.7085387.

[2] A. Awada, I. Viering, B. Wegmann, and A. Klein, "Application of Game Theory for Load Balancing in Long Term Evolution Networks," *Frequenz*, vol. 64, no. 9–10, pp. 180–184, Oct. 2010, https://doi.org/10.1515/FREQ.2010.64.9-10.180.

[3] H. H. Lu *et al.*, "A combined fibre/free-space-optical communication system for long-haul wireline/wireless transmission at millimetre-wave/sub-THz frequencies," *Communications Engineering*, vol. 2, no. 1, pp. 1–8, May 2023, https://doi.org/10.1038/s44172-023-00068-1.

[4] G. Pandey, A. Choudhary, and A. Dixit, "Wavelength Division Multiplexed Radio Over Fiber Links for 5G Fronthaul Networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 9, pp. 2789–2803, Sep. 2021, https://doi.org/10.1109/JSAC.2021.3064654.

[5] K. Singh, A. Dixit, and V. K. Jain, "Hybrid Medium Access Control Protocol for Radio-Over-Fiber Networks," *IEEE Access*, vol. 9, pp. 110889–110903, 2021, https://doi.org/10.1109/ACCESS.2021.3103513.

[6] B. G. Kim, S. H. Bae, H. Kim, and Y. C. Chung, "Feasibility of RoF-based optical fronthaul network for next-generation mobile communications," in *2017 Opto-Electronics and Communications Conference (OECC) and Photonics Global Conference (PGC)*, Singapore, Jul. 2017, pp. 1–2, https://doi.org/10.1109/OECC.2017.8114980.

[7] A. Udalcovs *et al.*, "An Insight into the Total Cost of Ownership of 5G Fronthauling," in *2018 20th International Conference on Transparent Optical Networks (ICTON)*, Bucharest, Romania, Jul. 2018, pp. 1–5, https://doi.org/10.1109/ICTON.2018.8474008.

[8] B. G. Kim, S. H. Bae, H. Kim, and Y. C. Chung, "Optical fronthaul technologies for next-generation mobile communications," in *2016 18th International Conference on Transparent Optical Networks (ICTON)*, Trento, Italy, Jul. 2016, pp. 1–3, https://doi.org/10.1109/ICTON.2016.7550604.

[9] S. Saibharath, S. Mishra, and C. Hota, "Swap-Based Load Balancing for Fairness in Radio Access Networks," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2412–2416, Aug. 2021, https://doi.org/10.1109/LWC.2021.3101983.

[10] W. K. Lai, Y. J. Yu, P. L. Tsai, and M. H. Shen, "A Multidepth Load-Balance Scheme for Clusters of Congested Cells in Ultradense Cellular Networks," *Wireless Communications and Mobile Computing*, vol. 2020, no. 1, 2020, Art. no. 6975890, https://doi.org/10.1155/2020/6975890.

[11] K. Suresh *et al.*, "Enhanced Metaheuristic Algorithm-Based Load Balancing in a 5G Cloud Radio Access Network," *Electronics*, vol. 11, no. 21, Jan. 2022, Art. no. 3611, https://doi.org/10.3390/electronics11213611.

[12] D. Wu *et al.*, "Reinforcement learning for communication load balancing: approaches and challenges," *Frontiers in Computer Science*, vol. 5, May 2023, https://doi.org/10.3389/fcomp.2023.1156064.

[13] B. Lim and M. Vu, "Distributed Multi-Agent Deep Q-Learning for Load Balancing User Association in Dense Networks," *IEEE Wireless Communications Letters*, vol. 12, no. 7, pp. 1120–1124, Jul. 2023, https://doi.org/10.1109/LWC.2023.3250492.

[14] R. Ahmad, M. D. Soltani, M. Safari, and A. Srivastava, "Reinforcement Learning-Based Near-Optimal Load Balancing for Heterogeneous LiFi WiFi Network," *IEEE Systems Journal*, vol. 16, no. 2, pp. 3084–3095, Jun. 2022, https://doi.org/10.1109/JSYST.2021.3088302.

[15] A. Lobinger, S. Stefanski, T. Jansen, and I. Balan, "Load Balancing in Downlink LTE Self-Optimizing Networks," in *2010 IEEE 71st Vehicular Technology Conference*, Taipei, Taiwan, 2010, pp. 1–5, https://doi.org/10.1109/VETECS.2010.5493656.

[16] H. M. Kavana, V. B. Kavya, B. Madhura, and N. Kamat, "Load Balancing using SDN Methodology," *International Journal of Engineering Research & Technology*, vol. 7, no. 5, pp. 206–208, May 2008.

[17] H. Wang, J. Ren, and T. Li, "Resource Allocation with Load Balancing for Cognitive Radio Networks," in *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*, Miami, FL, USA, Dec. 2010, pp. 1–5, https://doi.org/10.1109/GLOCOM.2010.5683966.

[18] T. Oyama and T. Seyama, "Load-Balancing Techniques for 5G Millimeter-Wave Distributed Antenna System," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, Helsinki, Finland, Apr. 2021, pp. 1–5, https://doi.org/10.1109/VTC2021-Spring51267.2021.9448953.

[19] "5G; NR; Physical layer procedures for data," ETSI, TS 138 214-V16.2.0-Release 16, 2020.

[20] D. Jain and B. Iyer, "Design and analysis of high-speed four-channel WDM Radio over Fiber system for Millimeter-wave applications," *International Journal of System Assurance Engineering and Management*, vol. 14, no. 3, pp. 746–758, Jul. 2023, https://doi.org/10.1007/s13198-020-01051-1.

[21] B. Tamrakar, K. Singh, and P. Kumar, "Performance Measurement of Radio over Fiber System at 20 GHz and 30 GHz by Employing with and without Optical Carrier Suppression," in *2019 6th International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, Mar. 2019, pp. 167–172.

[22] R. A. Shafik, Md. S. Rahman, and A. R. Islam, "On the Extended Relationships Among EVM, BER and SNR as Performance Metrics," in *2006 International Conference on Electrical and Computer Engineering*, Dhaka, Bangladesh, Dec. 2006, pp. 408–411, https://doi.org/10.1109/ICECE.2006.355657.

[23] T. T. H. Ly, H. S. Nguyen, T. S. Nguyen, V. V. Huynh, T. L. Nguyen, and M. Voznak, "Outage Probability Analysis in Relaying Cooperative Systems with NOMA Considering Power Splitting," *Symmetry*, vol. 11, no. 1, Jan. 2019, Art. no. 72, https://doi.org/10.3390/sym11010072.

[24] M. Alhabo, L. Zhang, and O. Oguejiofor, "Inbound handover interference-based margin for load balancing in heterogeneous networks," in *2017 International Symposium on Wireless Communication Systems (ISWCS)*, Bologna, Aug. 2017, pp. 146–151, https://doi.org/10.1109/ISWCS.2017.8108100.

[25] Y. Xu, W. Xu, Z. Wang, J. Lin, and S. Cui, "Load Balancing for Ultradense Networks: A Deep Reinforcement Learning-Based Approach," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9399–9412, Sep. 2019, https://doi.org/10.1109/JIOT.2019.2935010.

[26] G. Dalal, K. Dvijotham, M. Vecerik, T. Hester, C. Paduraru, and Y. Tassa, "Safe Exploration in Continuous Action Spaces." arXiv, Jan. 26, 2018, https://doi.org/10.48550/arXiv.1801.08757.

[27] J. Brownlee, "Understand the Impact of Learning Rate on Neural Network Performance," *MachineLearningMastery.com*, Jan. 24, 2019. https://www.machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/.

[28] P. E. I. Rivera and M. Erol-Kantarci, "Competitive Multi-Agent Load Balancing with Adaptive Policies in Wireless Networks." arXiv, Oct. 13, 2021, https://doi.org/10.48550/arXiv.2110.07050.

[29] "Keras documentation: Adam." https://keras.io/api/optimizers/adam/.

[30] M. Q. Abdulhasan, M. I. Salman, C. K. Ng, N. K. Noordin, S. J. Hashim, and F. Hashim, "An Adaptive Threshold Feedback Compression Scheme Based on Channel Quality Indicator (CQI) in Long Term Evolution (LTE) System," *Wireless Personal Communications*, vol. 82, no. 4, pp. 2323–2349, Jun. 2015, https://doi.org/10.1007/s11277-015-2350-1.

[31] "NORM.DIST Function," *Corporate Finance Institute*. https://corporatefinanceinstitute.com/resources/excel/norm-dist-function/.

[32] "NORMDIST function - Microsoft Support." https://support.microsoft.com/en-us/office/normdist-function-126db625-c53e-4591-9a22-c9ff422d6d58.

[33] A. Y. Al-Zahrani, "On the Statistical Distribution of Packets Service Time in Cellular Access Networks," *Engineering, Technology & Applied Science Research*, vol. 9, no. 6, pp. 4996–5000, Dec. 2019, https://doi.org/10.48084/etasr.3054.

[34] A. S. Afolabi, S. Ahmed, and O. A. Akinola, "A Reinforcement Learning Approach for Interference Management in Heterogeneous Wireless Networks," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 15, no. 12, Jun. 2021, Art. no. 65, https://doi.org/10.3991/ijim.v15i12.20751.