Kabarak Journal of Research & Innovation
www.kabarak.ac.ke

A Reinforcement Learning Approach to Service Based User Admission in a Multi-Tier 5G Wireless Networks

[1]Ojijo O. Mourice, [2]Ombuya O. Dismas, [3]Kipkebut Andrew, [4]Ayako Wycliffe & [5]Mochoge Cleophas

[1]mojijo@kabarak.ac.ke,[2]dombuya@kabarak.ac.ke,[3]akipkebut@kabarak.ac.ke, [4]wayako@kabarak.ac.ke, & [5]mochoge@kabarak.ac.ke

[1,2,3,4, &5]Kabarak University Dept. of Computer Science and Information Technology Nakuru, Kenya

Submitted 26th November, 2020, Accepted 24th August, 2021, Published 28th October, 2021

ABSTRACT

The massive connectivity in 5G wireless network is expected to become a challenge to communication network service providers. Many services over the 5G network will be aligned to a particular radio access technology (RAT). As a result, admitting a service-based user to a particular RAT will depend on the most efficient radio access technology selection (RAT). This is because 5G network will adopt multi-tier radio access networks ranging from high power macro base stations to extremely low power Bluetooth connectivity. Selection of a service-oriented RAT is critical because some wireless services have superior quality of service under certain RATs. Maintaining efficient RAT selection by network operators will improve power allocation efficiency, bandwidth allocation efficiency and operation expenditure. The complexity of associating a RAT to service-based user while considering network state such as data rate, the power allocation and hand-off frequency have not been fully explored. In this paper we propose a reinforcement learning approach to user admission based on efficient RAT selection considering wireless services in a cross tier wireless radio access network domain. The proposed algorithm indicates improved RAT selection efficiency considering transmit power, data rate and user hand off while minimizing the computation complexity. We perform extensive simulation using Python dynamic libraries and compare the finding against random association.

Key words—*Multi-Tier, RAT, Reinforcement Learning, 5G, and Wireless Networks*

1. INTRODUCTION

The fifth generation (5G) is expected to provide access to a multi-tier wireless access networks with many services being dedicated to certain radio access technologies for optimum performance. The complexity of determining the best RAT for a certain service is still a challenge and solving this problem requires an intelligent algorithm capable of achieving optimum performance. Furthermore, no single RAT can satisfy all the needs of all users based 5G services(Sandoval, Canovas-Carrasco, Garcia-Sanchez, & Garcia-Haro, 2019). The proposal of three main slices in 5G namely: ultra-reliable low latency communication (uRLLC), massive machine type communication (mMTC) and enhanced mobile broadband (eMBB) (Ojijo & Falowo, 2020)(Sunday O. Oladejo & Falowo, 2019; Sunday Oladayo Oladejo & Falowo, 2020)slices clearly depicts the need for service based RAT selection. The architectural design of 5G comprises multi-tier access ranging from a macro base station to a low power Bluetooth.

In this regard a user service will require an association with a specific RAT, for optimum performance. To reap the benefits of 5G network slicing and multi-tier design, a user can automatically be evaluated and connected to a specific base station(Xiang, Peng, Sun, & Yan, 2020). For instance, broadband access user may always be connected to WI-FI for optimal data rate and efficient power consumption and an internet of things user (IoT) may require a connection to a macro base station in order to provide connectivity to the million devices per kilometer feature to 5G mMTC slice. In this regard a reinforcement learning (RL) model for selecting a specific radio access-based user service if formulated for optimal association. Since the performance demand for specific services become stricter, enabling an efficient connectivity in 5G network reduces the cost of operation. The mathematical model of RL is an efficient method of intelligently allowing the radio access network (RAN) controller to associate a user to specific radio access technology(Sandoval et al., 2019)(Sun et al., 2018) based on the network characteristics and user demand. Furthermore, by allowing the RL agent to learn a specific policy π, the network user will be mapped to the optimal RAT based on a specific service request. Under this concept, we consider a user association to a macro base station, micro base station, picocell, femtocell, Wi-Fi, Bluetooth, and device to device (D2D).

The objective is to obtain a matching order for RAT that offers the best data rate under specific network conditions. In terms of services, we consider the subcategory services under the three known slices namely: multimedia, voice over internet protocol, internet of things, mission critical, and large file transfer. In the mentioned subcategories, both static and mobile users are considered. To achieve our mission, we build a finite state space containing all the possible user states. Whenever a mobile service is associated with a particular access technology, a weight is obtained matching how good the state is, this will be subsequently transformed into a reward function in the RL environment. The remainder of the paper is organized as follows. In section II we briefly describe different machine learning techniques III we provide a brief summary of the problem statement in our work. In section IV, the research objectives are clearly outlined. Section V provides a comparison with existing work from a variety of literature. In section VI we provide our system model. Section VII contains the simulation and results; we conclude our paper in section VIII where a brief summary of our work in the conclusion is outlined.

A. Machine Learning

According to Yu & He (2019), machine learning can be grouped into three categories namely:

- Supervised learning
- Unsupervised learning
- Reinforcement learning

Supervised learning (Zhang, Patras, & Haddadi, 2018) is a technique where the learning agent trains on a labeled data to construct a model for mapping an input data to the output data. Once the training is complete the model can be used to predict an output without relying on the training data. Examples of supervised learning models include decision tree, k-nearest neighbor, support vector machine, neural network, Bayes' theory, hidden Markov models and random forest.

Unsupervised Learning (Eugenio, Cayamcela, & Lim, 2018) is a model where input data are not labeled. The goal of the learning agent is to determine a common pattern with the unlabeled data by clustering the learned data into different groupings. Examples of unsupervised learning models are: self-organizing maps, and k-means.

Reinforcement Learning (Arulkumaran, Deisenroth, Brundage, & Bharath, 2017) provides a method of building a model that solves problems that require multi-stage decision making. It relies on Markov decision process to determine a policy for selecting an optimum action after visiting a state within an environment. Each action selection is rewarded. The goal of the learning agent is to maximize the reward obtained. The value of each action is stored as a Q-Value. All actions with maximum Q-Values are considered as the optimum actions. Examples of reinforcement leaning models include Q-learning, deep Q-learning, SARSA, dynamic programming, temporal difference and Monte-Carlo methods.

B. The Problem

The challenging act of determining which RAT to associate with a service is still an open challenge (Kildal, Vosoogh, & MacI, 2016). The economic aspect of reducing the operating expenditure is also an area of concern to many service providers of the 5G wireless network. Such liabilities arise from inefficient power consumption and bandwidth allocation due inefficient RAT selection. One way of reducing the cost of power consumption is efficient service based RAT selection (Xiang et al., 2020). Furthermore, the mathematical complexity of modeling a wireless network is highly intense; on that note a less complex scheme is widely accepted. Our approach promises a less complexity while maintaining a higher accuracy.

C. Objectives

In this work we intend to achieve the following objectives:

i. Model a reduced complexity environment considering user services for efficient RAT selection.
ii. Model a reinforcement learning environment considering user service requirements.
iii. Simulate and test a service-based RAT selection model in a finite space reinforcement learning environment.

II. LITERATURE REVIEW

Radio access technology selection has been previously studied by researchers; the authors interrogate some of the works already existing in comparison to the work in this paper. The work in (Sandoval et al., 2019) and (Sandoval, Canovas-Carrasco, Garcia-Sanchez, & Garcia-Haro, 2018) considers an IoT based RAT selection using RL. The experimental setting in this scenario is strictly based on static internet of things (IoT) devices and no consideration of mobile users was investigated as network condition generally degrades considerably as nodes become dynamic. Furthermore, IoT network resource allocation only belongs to the mMTC slice limiting the scope of broader investigation into other slices. In this paper, we consider a mobile user a paradigm not investigated in the reviewed paper.

The work in (Passas, Miliotis, Makris, & Korakis, 2019) investigated a distributed RAT selection considering multiple user applications. On a similar note, the author assumed static user environment. While this approach produced some interesting results, many 5G mobile user are non-static and must be considered for conclusive results. Further, the Lagrangian modeling require constraint relaxation for the problem solution which is highly mathematically intensive as compared to RL model where complex constraints can be part of the environment learned by agent without the need to solving the actual objective function.

The authors in (Anany, Elmesalawy, & El Din, 2019) proposed a two scenario RAT selection considering a long term evolution (LTE) and a wireless local area network (WLAN) using a matching game algorithm. While this approach produced some interesting results, the scope was very limited and does not reflect the current multi-tier network adopted in 5G. The study in (Ndashimye, Sarkar, & Ray, 2016) investigated a network selection mechanism for efficient handover. A fuzzy logic inference technique was used to enable seamless handover for vehicle to infrastructure network. The approach however is limited in its ability as it does not consider multitier and non-static or pedestrian users which in most cases form the majority users in a multitier network. The authors in (Perveen, Patwary, & Aneiba, 2019) presented a user admission control and slice allocation strategy where user characteristics such as data rate, bandwidth, priority, revenue to maximize user utility and resource limits. In this study, the authors did not consider a multi-tier environment which takes into consideration a realistic 5G environment. Authors in (Jiang, Condoluci, & Mahmoodi, 2016) proposed an inter and intra slice admission and resource allocation scheme and solved it using a heuristic approach. The study was based on both slice and user priority consideration. Slices and users with high priority were admitted considering resource constraints. This technique however did not consider a mobile user in multi-tier environments.

III. METHODOLOGY

In this research the authors consider a multi-tier 5G network consisting of the following: A macro-cell, a microcell, a pico-cell, a femtocell, a Wi-Fi cell, Bluetooth connectivity and device to device (D2D) connectivity. Assuming a user $\boldsymbol{u}\in \boldsymbol{U}$ under a cell $\boldsymbol{c}\in \boldsymbol{C}$ with down link power $\boldsymbol{P_c}$ and a maximum cell data rate $\boldsymbol{d_c}$ is connected to any of the cells in Fig. 1.

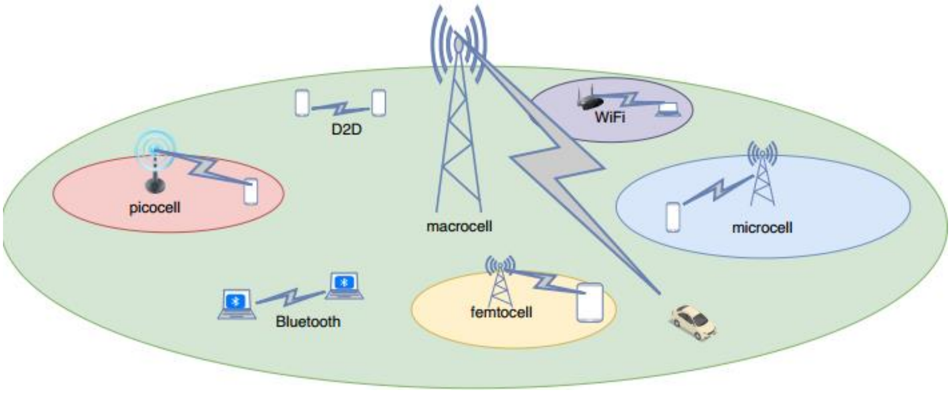


Fig. 1. 5G multi-tier RATs

A user $\boldsymbol{u}$ is assumed to access an application $\boldsymbol{g} \in \boldsymbol{G}$ requiring a minimum data rate of $\boldsymbol{r_g^u}$ in any cell $\boldsymbol{c}$. To meet the mobility constraint, the user can be assumed to have a mobility status ($\boldsymbol{m^u}\in \{\boldsymbol{0{:}1}\}$), where $\boldsymbol{0}$ implies a static user and $\boldsymbol{1}$ implies a mobile user. The study assume that a mobile user is one in a car at a constant velocity for simplicity reasons while a static user is one who is immobile or walking. Each cell is considered to have a radius $\boldsymbol{w_c}$. The handover (HO) rate can therefore be given by eq.(1) as:

$$\boldsymbol{HO_u = \frac{m^u}{w_c}} \qquad (1)$$

The cost incurred by a cell provider when user is given access can provided by eq. (2) as:

$$\boldsymbol{cost = P_c\, d_c} \qquad (2)$$

The price paid by any user accessing a cell $\boldsymbol{c}$ having a downlink data rate $\boldsymbol{r_g^u}$is given by eq.(3)

$$\boldsymbol{cost_u} = \frac{\boldsymbol{r_g^u}}{\boldsymbol{cost}} \times (\boldsymbol{HO_u})^{-1} \qquad (3)$$

The cost $\boldsymbol{cost_u}$ will eventually be mapped to the reward function in the RL. To achieve the optimal goal of this study, the aim is associate a user to a cell which offers maximum power, data rate and minimum number of handover. In this regard eq (4)-eq (6) denote optimum conditions for user to cell association.

$$\max_{\boldsymbol{P_c}} \boldsymbol{Pc}\ \forall\ \boldsymbol{c}\ \in \boldsymbol{C} \qquad (4)$$

$$\max_{\boldsymbol{r_g^u}} \boldsymbol{r_g^u}\ \ \forall\ \boldsymbol{u}\ \in \boldsymbol{U},\ \forall\ \boldsymbol{g}\ \in \boldsymbol{G} \qquad (5)$$

$$\min_{\boldsymbol{HO_u}} \boldsymbol{HO_u}\ \ \forall\ \boldsymbol{u}\ \in \boldsymbol{U} \qquad (6)$$

A. *Reinforcement Learning*

Any problem that can be formulated as a Markov-decision process (MDP) can be solved using reinforcement learning(Sutton & Barto, 2017). To select the most suitable RAT for a user $\boldsymbol{u}$ accessing an application $\boldsymbol{g}$, the mathematical framework required in RL is to map an action ($\boldsymbol{a} \in \boldsymbol{A}$) (selection a RAT) to the existing user state ($\boldsymbol{s} \in \boldsymbol{S}$) (network access conditions and user demands).Assuming a user has the ability to connect to any of the RAT at any time considering the mobility status $\boldsymbol{m}$, then the reward ($\boldsymbol{R|s{:}a}$) obtained for an action (RAT selection) is given by

$$\boldsymbol{R} = \frac{\boldsymbol{\beta}}{\boldsymbol{cost_u} - \boldsymbol{V_a}(\boldsymbol{s}) + \Delta} \qquad (7)$$

where $\boldsymbol{\beta}$ is a constant selected to regulate the value of the reward function, $\boldsymbol{V_a}(\boldsymbol{s})$ is the action value after observing any state $\boldsymbol{s}$ and Δ is a small value ensuring the denominator does not become zero. The action value $\boldsymbol{V_a}(\boldsymbol{s})$ is denoted by

$$\boldsymbol{V_a}(\boldsymbol{s}) = \boldsymbol{cost}\ \times\ \boldsymbol{a} \qquad (8)$$

where $\boldsymbol{a}$ denotes the action taken.

B. *State Space*

The novel state space employed in this study comprises a three-dimensional tensor given by

$$\boldsymbol{S_{i,j}^n} = [\boldsymbol{w_{i,j}^n}, \boldsymbol{m_{i,j}^n}, \boldsymbol{P_{i,j}^n}, \boldsymbol{r_{i,j}^n}] \qquad (9)$$

where $\boldsymbol{w_{i,j}^n}, \boldsymbol{m_{i,j}^n}, \boldsymbol{P_{i,j}^n}, \boldsymbol{r_{i,j}^n}$ is the cell radius, user mobility status, cell transmit power and user data rate at a time instant $\boldsymbol{n}$ in an $\boldsymbol{i} \times \boldsymbol{j}$ tensor. At any time, instant $\boldsymbol{n}$ which is equivalent to a particular unit of time in an episode the agent observes the state space which is a $\boldsymbol{i} \times \boldsymbol{j}$ tensor constituting $\boldsymbol{i}$ rows and $\boldsymbol{j}$ columns. In the Q-learning algorithm, the index $\boldsymbol{k}$ is equivalent to the row index $\boldsymbol{i}$ in the tensor. In each iteration the agent runs through all the $\boldsymbol{j} = \boldsymbol{J}$ elements in row $\boldsymbol{i}$.

C. *Action space*

The action space constitutes the set of all actions the agent will take during the learning process. To reach optimality the agent must learn to take an optimum action after each state observation. The action space is given by the tuple $\boldsymbol{a_i} = [\boldsymbol{a_i^{micro}}, \boldsymbol{a_i^{macro}}, \boldsymbol{a_i^{pico}}, \boldsymbol{a_i^{D2D}}, \boldsymbol{a_i^{Wifi}}, \boldsymbol{a_i^{femto}}, \boldsymbol{a_i^{bt}}]$, where $\boldsymbol{a_i^{micro}}$

denotes microcell selection, a_i^{macro} macrocell selection, a_i^{pico} pico-cell selection, a_i^{D2D} device to device selection, a_i^{Wifi} Wi-Fi selection, a_i^{femto} femtocell selection and a_i^{bt} Bluetooth network selection. Given a specific state, the agent

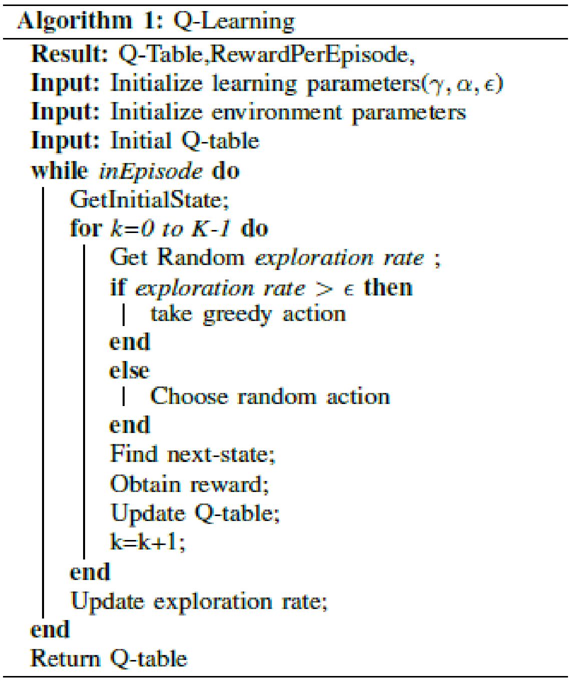
takes an action, an optimum action is when a user is associated to a cell with maximum power, maximum data rate and lowest handover rate.

D. Q-Learning

To achieve the objective of determining the optimum actions, the study employ the Q-learning method to evaluate the value of each action considering individual states. The action-state pair with highest Q-value becomes the optimal solution. The iteration to evaluate each action ends up with the update in the Q-table where each action is mapped to a corresponding state. The **Algorithm 1** Q-learning employs the classic Bellman's equation given by eq. (10) (Arulkumaran et al., 2017)(Bega, Costa-Perez, Gramaglia, Sciancalepore, & Banchs, 2019)(Bega et al., 2020)(Bega, Gramaglia, et al., 2019)(Sun et al., 2018)

$$Q(s,a) = (1-\alpha)q(s,a) + \alpha\{R_{t+1} + \gamma \max_{a'} q(a',s')\} \qquad (10)$$

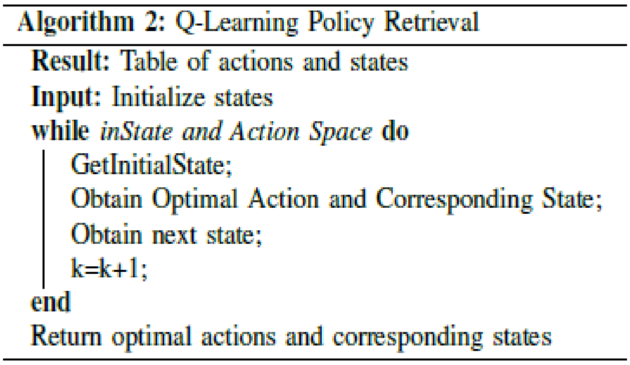
where α is the learning rate, γ is a discount factor and Rt is the long-term reward observed at time t. The term $q(a,s)$ represents the previous q-value, $q(a',s')$ is the maximum Q-value in the Q-table. During the learning process, the Q-table is updated by eq. (10) until all episodes are completed. In general, the maximum Q-values per row are obtained and mapped to the corresponding states and actions during the policy retrieval in **Algorithm 2**. Once this is completed the end result is policy table with only states and optimum actions.



```
Algorithm 1: Q-Learning
Result: Q-Table,RewardPerEpisode,
Input: Initialize learning parameters($\gamma, \alpha, \epsilon$)
Input: Initialize environment parameters
Input: Initial Q-table
while inEpisode do
    GetInitialState;
    for k=0 to K-1 do
        Get Random exploration rate ;
        if exploration rate > $\epsilon$ then
            take greedy action
        end
        else
            Choose random action
        end
        Find next-state;
        Obtain reward;
        Update Q-table;
        k=k+1;
    end
    Update exploration rate;
end
Return Q-table
```

A. Policy Retrieval

The algorithm in **Algorithm 2** is how the policy is retrieved



```
Algorithm 2: Q-Learning Policy Retrieval
Result: Table of actions and states
Input: Initialize states
while inState and Action Space do
    GetInitialState;
    Obtain Optimal Action and Corresponding State;
    Obtain next state;
    k=k+1;
end
Return optimal actions and corresponding states
```

IV. SIMULATIONSAND RESULTS

Table 1. Simulation parameters

Hyper parameters	**Value**
Learning rate α	0.001
Discount rate γ	0.09
Number of episodes	5000
Minimum exploration rate	0.01
Exploration decay rate	0.001
Other parameters	**Value**
Maximum cell radius	5km
Minimum cell radius	1m
Maximum data rate	1Gbps
Minimum data rate	15Mbps
Maximum transmit power	58.5dBm
Minimum transmit power	33dBm

In our simulation the study considered a seven tier 5G cell network each of specific maximum cell power ranging from a minimum of 33dBm in a D2D architecture to 58.5dBm in a macrocell. The study also considers a minimum cell data rate of 15 mbps in a D2D network to a maximum of 1 giga bit per second (GBPS) possible in 5G a microcell. The radii were chosen to range from 4m to a maximum of 10km in a macro-cell. Each user may have a mobility status of mobile or static. After 5000 episodes of learning the results of simulations were found as follows. In Fig.2 the study present rewards per episode where the agent was observed to converge to the optimal solution after 2500 episodes.

The evaluated policy is represented in Fig. 3. The study paired each user state to the selected RAT. Each red spot in the graph represents the pairing region while the remaining blue areas represent no pairing. In summary, it can be observed that the agent ignored pairing D2D and Bluetooth cells to any user due to the initialized low data rate while pairing a maximum of 4 user states to a microcell considered having the highest data rate. In Fig 4 the study compared the allocated power to any paired user with the cell radius. The proposed algorithm continually increments the cell power as the cell radius increased which is the standard practice in cellular networks. In random power allocation the study evaluated has lower efficiency; this can be observed as small cells were allocated high power. In **Fig 5**. the study evaluated the data rate considering the cell radius. The tendency was that the rate dropped as the user moved away from the base station, except in a microcell considered to have the highest data rate. The random data rate allocation once again did not perform efficient data rate allocation considering the inconsistent fluctuations.

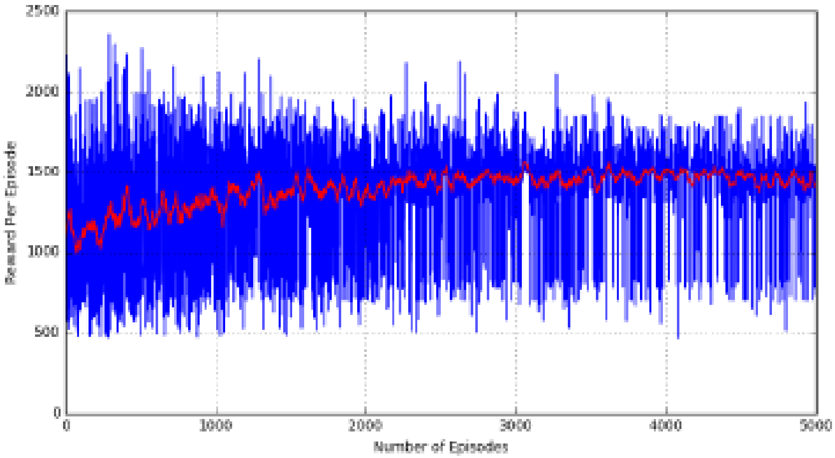
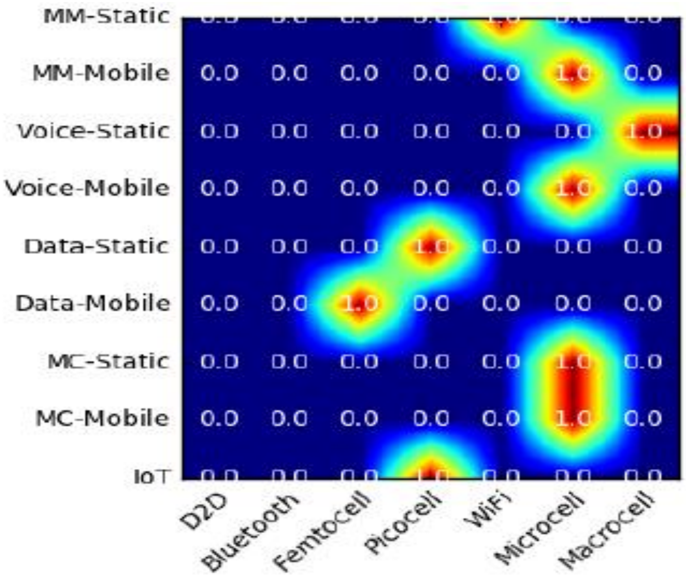


Fig. 2. Reinforcement learning showing rewards per episode



	D2D	Bluetooth	Femtocell	Picocell	WiFi	Microcell	Macrocell
MM-Static	0.0	0.0	0.0	0.0	1.0	0.0	0.0
MM-Mobile	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Voice-Static	0.0	0.0	0.0	0.0	0.0	0.0	1.0
Voice-Mobile	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Data-Static	0.0	0.0	0.0	1.0	0.0	0.0	0.0
Data-Mobile	0.0	0.0	1.0	0.0	0.0	0.0	0.0
MC-Static	0.0	0.0	0.0	0.0	0.0	1.0	0.0
MC-Mobile	0.0	0.0	0.0	0.0	0.0	1.0	0.0
IoT	0.0	0.0	0.0	1.0	0.0	0.0	0.0

Fig. 3. User state vs RAT selection

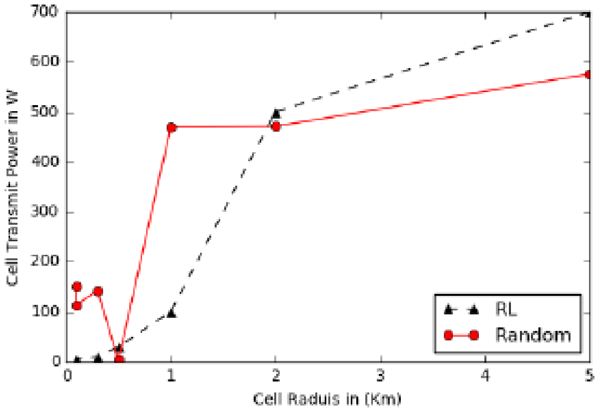


Fig. 4. Radiated cell power selection vs cell radius

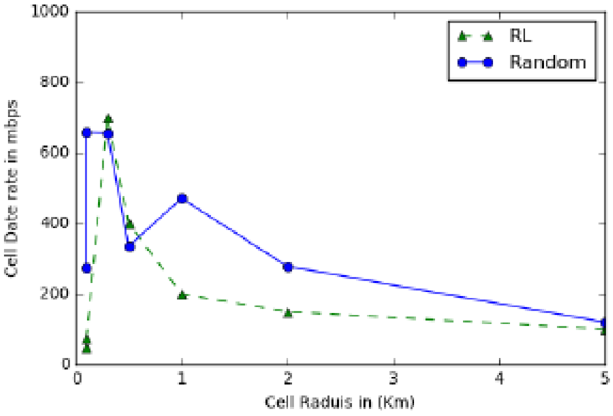


Fig. 5. Allocated cell data rate vs cell radius

V. CONCLUSION AND RECOMMENDATIONS

In conclusion, the study has presented an RL based RAT selection scheme for specific use case. The study shows that the proposed technique has improved efficiency in associating a user to RAT considered the required cell power and data rates. The study presents a novel RL environment and reward function for state-action evaluation. The study presented results compared to the random association, it is observed that the mechanism out performs the random mechanism. The consideration of a finite state space is however a limitation in this study as state spaces may become continuous leading inefficient memory use if Q-learning is considered under what is known as the curse of dimensionality. It therefore recommended that for continuous and large state spaces, a deep Q-learning approach be considered. In this regard function approximators such as neural networks be employed for better performance.

VI. REFERENCES

Anany, M. G., Elmesalawy, M. M., & El Din, E. S. (2019). A Matching Game Solution for Optimal RAT Selection in 5G Multi-RAT HetNets. *2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2019*, 1022–1028. https://doi.org/10.1109/UEMCON47517.2019.8993013

Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, *34*(6), 26–38. https://doi.org/10.1109/MSP.2017.2743240

Bega, D., Costa-Perez, X., Gramaglia, M., Sciancalepore, V., & Banchs, A. (2019). A Machine Learning approach to 5G Infrastructure Market optimization. *IEEE Transactions on Mobile Computing*, *XX*(X), 1–1. https://doi.org/10.1109/tmc.2019.2896950

Bega, D., Gramaglia, M., Fiore, M., Banchs, A., Costa-Perez, X., & Costa-Perez DeepCog, X. (2019). *Network Management in Sliced 5G Networks with Deep Learning*. *IEEE INFOCOM*. Retrieved from https://hal.inria.fr/hal-01987878

Bega, D., Gramaglia, M., Garcia-Saavedra, A., Fiore, M., Banchs, A., & Costa-Perez, X. (2020). Network Slicing Meets Artificial Intelligence: An AI-Based Framework for Slice Management. *IEEE Communications Magazine*, *58*(6), 32–38. https://doi.org/10.1109/MCOM.001.1900653

Eugenio, M., Cayamcela, M., & Lim, W. (2018). Artificial Intelligence in 5G Technology : A Survey. *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, (Ml), 860–865. https://doi.org/10.1109/ICTC.2018.8539642

Jiang, M., Condoluci, M., & Mahmoodi, T. (2016). Network slicing management & prioritization in 5G mobile systems. *European Wireless 2016; 22th European Wireless Conference*, 197–202. Retrieved from http://ieeexplore.ieee.org/document/7499297/

Kildal, P. S., Vosoogh, A., & MacI, S. (2016). Fundamental Directivity Limitations of Dense Array Antennas: A Numerical Study Using Hannan’s Embedded Element Efficiency. *IEEE Antennas and Wireless Propagation Letters*, *15*, 766–769. https://doi.org/10.1109/LAWP.2015.2473136

Ndashimye, E., Sarkar, N. I., & Ray, S. K. (2016). A Novel Network Selection Mechanism for Vehicle-to-Infrastructure Communication. *Proceedings - 2016 IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, DASC 2016, 2016 IEEE 14th International Conference on Pervasive Intelligence and Computing, PICom 2016, 2016 IEEE 2nd International Conference on Big Data*, 483–488. https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2016.94

Ojijo, M. O., & Falowo, O. E. (2020). A Survey on Slice Admission Control Strategies and Optimization Schemes in 5G Network. *IEEE Access*, *8*, 14977–14990.

Oladejo, Sunday O., & Falowo, O. E. (2019). Latency-Aware Dynamic Resource Allocation Scheme for 5G Heterogeneous Network: A Network Slicing-Multitenancy Scenario. In *International Conference on Wireless and Mobile Computing, Networking and Communications* (Vol. 2019-Octob, pp. 1–7). IEEE. https://doi.org/10.1109/WiMOB.2019.8923397

Oladejo, Sunday Oladayo, & Falowo, O. E. (2020). Latency-Aware Dynamic Resource Allocation Scheme for Multi-Tier 5G Network: A Network Slicing-Multitenancy Scenario. *IEEE Access*, *8*(2), 74834–74852. https://doi.org/10.1109/ACCESS.2020.2988710

Passas, V., Miliotis, V., Makris, N., & Korakis, T. (2019). Dynamic RAT Selection and Pricing for Efficient Traffic Allocation in 5G HetNets. *IEEE International Conference on Communications*, *2019-May*. https://doi.org/10.1109/ICC.2019.8761831

Perveen, A., Patwary, M., & Aneiba, A. (2019). Dynamically Reconfigurable Slice Allocation and Admission Control within 5G Wireless Networks. *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, 1–7. https://doi.org/10.1109/vtcspring.2019.8746625

Sandoval, R. M., Canovas-Carrasco, S., Garcia-Sanchez, A. J., & Garcia-Haro, J. (2018). Smart usage of multiple rat in IoT-oriented 5G networks: A reinforcement learning approach. *10th ITU Academic Conference Kaleidoscope: Machine Learning for a 5G Future, ITU K 2018*, 1–8. https://doi.org/10.23919/ITU-WT.2018.8597940

Sandoval, R. M., Canovas-Carrasco, S., Garcia-Sanchez, A. J., & Garcia-Haro, J. (2019). A reinforcement learning-based framework for the exploitation of multiple rats in the iot. *IEEE Access*, *7*, 123341–123354. https://doi.org/10.1109/ACCESS.2019.2938084

Sun, Q., I, C.-L., Zhao, Z., Zhang, H., Chen, X., Yang, C., … Zhao, M. (2018). Deep Reinforcement Learning for Resource Management in Network Slicing. *IEEE Access*, *6*, 74429–74441. https://doi.org/10.1109/access.2018.2881964

Sutton, R. S., & Barto, A. G. (2017). *Reinforcement learning Complete Draft*. *https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf*. Retrieved from https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf

Xiang, H., Peng, M., Sun, Y., & Yan, S. (2020). Mode Selection and Resource Allocation in Sliced Fog Radio

Access Networks: A Reinforcement Learning Approach. *IEEE Transactions on Vehicular Technology*, *69*(4), 4271–4284. https://doi.org/10.1109/TVT.2020.2972999

Yu, F. R., & He, Y. (2019). Deep Reinforcement Learning for Interference Alignment Wireless Networks. In *SpringerBriefs in Electrical and Computer Engineering. Springer, Cham* (pp. 21–44). Springer, Cham. https://doi.org/https://doi.org/10.1007/978-3-030-10546-4_3

Zhang, C., Patras, P., & Haddadi, H. (2018). Deep Learning in Mobile and Wireless Networking: A Survey, 1–67. Retrieved from http://arxiv.org/abs/1803.04311