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

1st Ojijo Mourice. O
*Kabarak University*
*Dept. of Computer Science and Information Technology*
Nakuru, Kenya
mojijo@kabarak.ac.ke

2nd Ombuya Dismas. O
*Kabarak University*
*Dept of Computer Science and Information Technology)*
Nakuru, Kenya
dombuya@kabarak.ac.ke

3rd Kipkebut. Andrew
*Kabarak University*
*Dept of Computer Science and Information Technology*
Nakuru, Kenya
akipkebut@kabarak.ac.ke

4th Ayako Wyclife.
*Kabarak University*
*Dept of Computer Science and Information Technology*
Nakuru, Kenya
wayako@kabarak.ac.ke

5th Mochoge Cleophas.
*Kabarak University*
*Dept of Computer Science and Information Technology*
Nakuru, Kenya
mochoge@kabarak.ac.ke

***Abstract*—The expected massive connectivity in 5G wireless network is bound to become a challenge to service providers. Many services over the 5G network will be aligned to a particular radio access network (RAN). As a result admitting a service based user to a particular RAN 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 service packet size, the turn around time, the power allocation has not to 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 is expected to improve RAT selection efficiency while minimizing the computation complexity. We perform extensive simulation using Python dynamic libraries and present our results along side existing approaches.**

***Index Terms*—Multi-Tier, RAT, RAN Reinforcement Learning, 5G, Wireless Networks**

## I. INTRODUCTION

The fifth generation (5G) is expected to provide access to a multi-tier wireless with many services being dedicated to certain radio access technologies for optimum performance. The complexity of determining which RAT is best 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 user based 5G services [1]. 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) [2] 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 based 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 [3]. For instance, broadband access user may always be connected to WI-FI for optimal data rate and efficient power consumption. In this regard we formulate a reinforcement learning (RL) model for selecting a specific radio access based user service. Since the performance demand for specific services become more strict, enabling an efficient connectivity and 5G network reduces the cost of operation.

The mathematical model of RL is an efficient method of intelligently allowing the RAN controller to associate a user to specific radio access technology. Furthermore, by allowing the RL agent to learn a specific policy $\pi$, the network user

Identify applicable funding agency here. If none, delete this.

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(mobile and static), voice over internet protocol(mobile and static), internet of things, mission-critical(mobile and static), and file transfer(mobile and static).

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 provide a brief summary of the problem statement in our work. In section III, the research objectives are clearly outlined. Section IV provides a comparison with existing work from a variety of literature. In section V we provide our system model. Section VI contains the simulation and results, we conclude our paper in section VII where a brief summary of our work in the conclusion is outlined.

## II. The Problem

The challenging act of determining which RAT to associate with a service is still an open challenge [4]. 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 [3]. 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 complex scenario while maintaining a high accuracy level.

## III. Objectives

In this work we intend to achieve the following objectives.

- Model a reduced complexity environment considering user services for efficient RAT selection.
- Model a reinforcement learning environment considering the a dynamic user service requirement.
- Simulate and test a service based RAT selection model in a finite space reinforcement learning environment.

## IV. Literature Review

RAT selection has been previously studied by researches, we interrogate some of the works already existing in comparison to the work in this paper. The work in [1] and [5] 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 is investigated as network condition generally degrades considerably as a node becomes dynamic. Furthermore IoT network resource allocation belongs to the mMTC slice limiting the scope of broader investigation into other slices. In this paper we intend to consider the possibility of having a mobile user.

Passas *et al* in [6] investigate 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.

## V. System Model

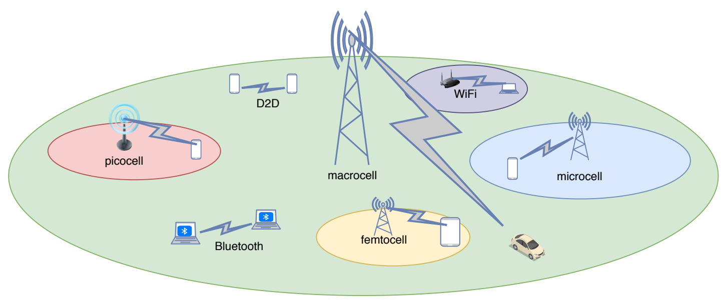


Fig. 1. 5G multi-tier RATs

We consider a multi-tier 5G network consisting of the following: A macrocell ,a microcell, a picocell,a femtocell, a WiFi cell, Bluetooth connectivity and device to device (D2D) connectivity. Assuming a user $u$ under a cell $c$ with downlink power $P_c$ and a cell data rate $d_c$ is connected to any of the cells in Fig. 1. Also the user can be assumed to be accessing an application $g$ requiring a minimum data rate of $r_g^u$ under any cell $c$. To meet the mobility constraint, the user can be assumed to have a mobility status ( $m^u \in 0, 1)$, where $0$ implies a static user and $1$ implies a mobile user. We 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 $k_c$. The handover (HO) rate can therefore be given by eq.(1) as:

$$HO = \frac{m^u}{k_c} \tag{1}$$

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

$$cost = P_c \times d_c \tag{2}$$

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

$$cost_u = \frac{r_g^u}{cost} \times \frac{1}{HO} \tag{3}$$

The cost $cost_u$ will eventually be mapped to the reward function in the RL.

## VI. Reinforcement Learning

Any problem that can be formulated as a multi-decision process (MDP) can be solved using reinforcement learning [7]. To select the most suitable RAT for a user $u$ accessing an application $g$, the mathematical framework required in RL is to map an **action** ($a \in A$) (selection a RAT) to the existing user **state** ($s \in S$) (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 $m$, then the reward $(\mathfrak{R}|s, a)$ obtained for an action (RAT selection) is given by

$$\mathfrak{R} = \frac{\beta}{cost_u - V_a(s) + \Delta} \tag{4}$$

where $\beta$ is a constant selected to regulate the value of the reward function ,$V_a(s)$ is the value the action taken under state $s$ and $\Delta$ a small value ensuring the denominator does not become zero.

### A. Q-Learning

To achieve the objective of determining the optimum actions we 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. 5 [8] [9]

$$Q(\boldsymbol{s},\boldsymbol{a}) = (1-\alpha)q(\boldsymbol{s},\boldsymbol{a}) + \alpha\{R_{t+1} + \gamma \max_{a'} q(\boldsymbol{a'},\boldsymbol{s'})\} \tag{5}$$

where $\alpha$ is the learning rate, $\gamma$ is a discount factor and $R_t$ is the long term reward observed at time $t$.

The end result in the Q-Table is optimal Q-values, this is however not complete until the policy is obtained under what is know as policy retrieval. The algorithm in Algorithm 2 is how the policy is retrieved.

## VII. Results and Simulations

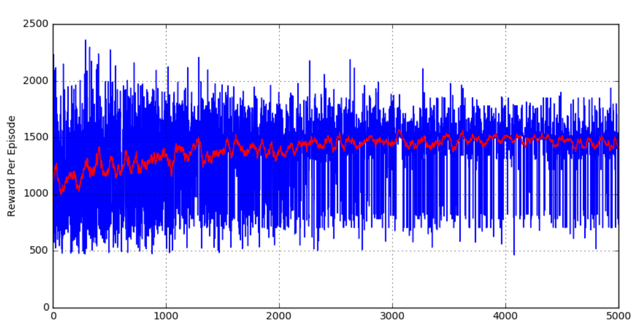


Fig. 2. Reinforcement learning showing rewards per episode

In our simulation we considered a seven tier 5G cell network each of specific maximum cell power ranging from a minimum

**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

**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

of 2 watts in a D2D architecture to 700 watts in a macrocell. We also consider a minimum cell data rate of 15 mbps in a D2D network to a maximum of 1 gbps possible in 5G a microcell. The radii were chosen to range from 4m to a maximum of 10km in a macrocell. Each user may have a mobility status of mobile or static. The data rate per user was varied based on the user cell association and selection beside the mobility status.

After 5000 episodes of learning the results of simulations were found as follows. In Fig.2 we 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. We paired each user state to the selected RAT. Each red spot in the graph represent the pairing region while the remaining blue areas represents no pairing.

In summary, it can be observed that the agent ignored pairing D2D and Bluetooth cells to any user due to their initialized low data rate while pairing a maximum of 4 user

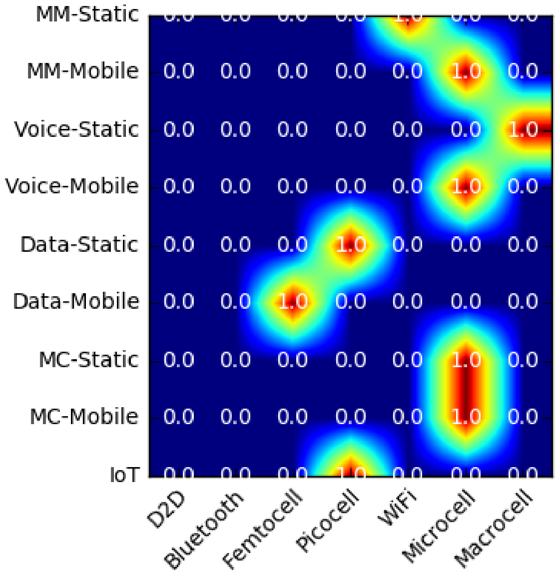


Fig. 3. User state vs RAT selection

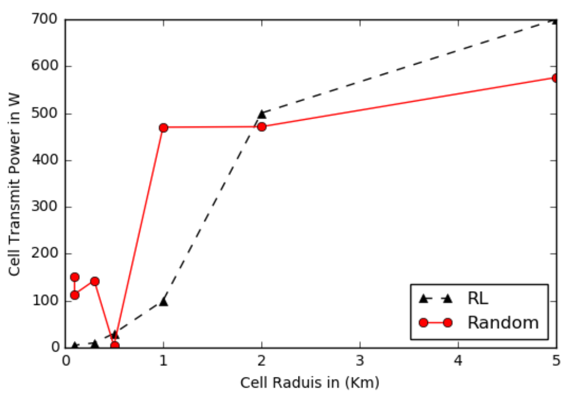


Fig. 4. Radiated cell power selection vs cell radius

states to a microcell considered to have the highest data rate. In Fig 4 we compared the allocated power to any paired user with the cell radius. The proposed algorithm continually increment the cell power as the cell radius increased which is the standard practice in cellular networks. The random power allocation we evaluated has lower efficiency, this can be observed as small cells were allocated high power.

In Fig 5 we evaluated the data rate considering the cell radius. The rate general tendency was to drop as the user moved a 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.

## VIII. Conclusion

In conclusion, we have presented an RL based RAT selection mechanism based on user application. We have shown that

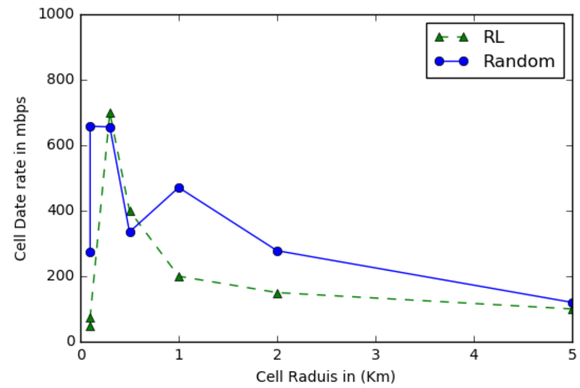


Fig. 5. Allocated cell data rate vs cell radius

our proposed technique has a high efficiency in associating a user to RAT considered the required cell power and data rates. We build an RL environment considering a novel reward function for state-action evaluation. We presented our results compared to the random association, we observed the our mechanism out performs the random mechanism.

The consideration of a finite state space is how ever a limitation in our work as most states are continuous leading inefficient memory use if Q-learning is considered. In our future work we will consider function approximation in deep neuron networks and deep Q-learning.

## References

[1] R. M. Sandoval, S. Canovas-Carrasco, A. J. Garcia-Sanchez, and J. Garcia-Haro, "A reinforcement learning-based framework for the exploitation of multiple rats in the iot," *IEEE Access*, vol. 7, pp. 123 341–123 354, 2019.

[2] M. O. Ojijo and O. E. Falowo, "A Survey on Slice Admission Control Strategies and Optimization Schemes in 5G Network," *IEEE Access*, vol. 8, pp. 14 977–14 990, 2020.

[3] H. Xiang, M. Peng, Y. Sun, and S. Yan, "Mode Selection and Resource Allocation in Sliced Fog Radio Access Networks: A Reinforcement Learning Approach," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4271–4284, 2020.

[4] P. S. Kildal, A. Vosoogh, and S. MacI, "Fundamental Directivity Limitations of Dense Array Antennas: A Numerical Study Using Hannan's Embedded Element Efficiency," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 766–769, 2016.

[5] R. M. Sandoval, S. Canovas-Carrasco, A. Garcia-Sanchez, and J. Garcia-Haro, "Smart usage of multiple rat in iot-oriented 5g networks: A reinforcement learning approach," in *2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K)*, 2018, pp. 1–8.

[6] V. Passas, V. Miliotis, N. Makris, and T. Korakis, "Dynamic RAT Selection and Pricing for Efficient Traffic Allocation in 5G HetNets," *IEEE International Conference on Communications*, vol. 2019-May, 2019.

[7] R. S. Sutton and A. G. Barto, "Reinforcement learning Complete Draft," 2017.

[8] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[9] D. Bega, X. Costa-Perez, M. Gramaglia, V. Sciancalepore, and A. Banchs, "A Machine Learning approach to 5G Infrastructure Market optimization," *IEEE Transactions on Mobile Computing*, vol. XX, no. X, pp. 1–1, 2019.

[10] M. G. Anany, M. M. Elmesalawy, and E. S. El Din, "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*, pp. 1022–1028, 2019.

[11] E. Ndashimye, N. I. Sarkar, and S. K. Ray, "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*, pp. 483–488, 2016.

[12] Q. Sun, C.-L. I, Z. Zhao, H. Zhang, X. Chen, C. Yang, R. Li, and M. Zhao, "Deep Reinforcement Learning for Resource Management in Network Slicing," *IEEE Access*, vol. 6, pp. 74 429–74 441, 2018.