Reinforcement Learning Algorithm for Traffic Steering in Heterogeneous Network

Cezary Adamczyk
Institute of Radiocommunications
Poznan University of Technology
Poznan, Poland
Email: cezary.adamczyk@student.put.poznan.pl

Adrian Kliks
Institute of Radiocommunications
Poznan University of Technology
Poznan, Poland
Email: adrian.kliks@put.poznan.pl

***Abstract*—Heterogeneous radio access networks require efficient traffic steering methods to reach near-optimal results in order to maximize network capacity. This paper aims to propose a novel traffic steering algorithm for usage in HetNets, which utilizes a reinforcement learning algorithm in combination with an artificial neural network to maximize total user satisfaction in the simulated cellular network. The novel algorithm was compared with two reference algorithms using network simulation results. The results prove that the novel algorithm provides noticeably better efficiency in comparison with reference algorithms, especially in terms of the number of served users with limited frequency resources of the radio access network.**

***Index Terms*—traffic steering, heterogeneous network, reinforcement learning, neural network**

I. Introduction

Evolution of telecommunication technologies allows cellular network operators to use radio resources more efficiently and therefore improve service quality and serve more users. Introduction of each new generation of radio access network architecture and gradual implementation of the new technology leads to a situation where several radio access technologies coexist creating a heterogeneous network (HetNet) [1]. Additional heterogeneity in network architecture is caused by many base stations types being installed for different coverage scenarios. These include macro cells for large area coverage and small cells (i.e. micro-, pico-, femtocells) for more throughput-demanding coverage or areas not covered by macro cells [2].

Such complex and diverse RAN architecture causes many challenges for the network operator, including backhaul provision for each base station, inter-RAT and intra-RAT interference mitigation and fine-tuning base station parameters to maximize its capacity [3]. One of many important aspects of HetNets optimization is traffic steering, i.e. load balancing [4]. It is the process of adaptive traffic allocation to different base stations (or in more generic approach - available radio resources), often with a certain priority set by the network operator, e.g. to maximize total network throughput.

As in heterogeneous networks (HetNets) one user can often be in range of many base stations, the load balancing algorithm must decide on which cell's radio resources should be allocated to the user. Basic traffic steering algorithms rely on a single criterion to make a decision on what cell should serve given user. This criterion may be current load percentages of in-range base stations or satisfaction of user's bit rate demand with radio resources available to be allocated by each base station. These approaches guarantee moderate efficiency at a low computational cost, but are far from optimal.

The key paper contribution of this paper is a proposal of an universal traffic steering algorithm for utilization in HetNets. The solution includes a dedicated data-processing flow that combines ANN inference and SARSA algorithm for ANN optimization to provide near-optimal traffic steering. The idea itself is inspired by work described in [5], but aims to provide more universal utility for all generations of radio access network, including 5G and beyond.

II. Traffic steering methods

A. Problem statement and reference scenario

The key problem addressed in this paper deals with traffic steering in HetNet scenarios. In general, the goal of traffic steering algorithms is to

978-1-6654-2854-5/21/$31.00 ©2021 IEEE

serve users using radio access network's resources in a way that uses the resources most optimally in regard to a chosen criterion. Basic approaches mentioned in [5] include allocating radio resources of the least loaded cell (denoted hereafter as Classic Load Balancing, *CLB*) or allocating radio resources of the cell that provides best user satisfaction (Satisfaction-based Load Balancing, *SLB*). In our approach, we aim to assign users to the cells in order to maximize total user satisfaction.

B. Reinforcement Learning Load Balancing

In the proposed scheme, we steer the cellular traffic within the HetNet by utilizing a reinforcement learning scheme called SARSA for optimization of an ANN [6]. In turn, the ANN adapts its model to use the limited radio resources in a way that maximizes the total user satisfaction in given scenario. The name of the SARSA method comes from the components used to update the state-action value function estimate [6]:

- S_t - environment state observation in *t*
- A_t - action taken in *t*
- R_{t+1} - reward received in *t*+1 after taking action A_t
- S_{t+1} - environment state observation in *t*+1
- A_{t+1} - action taken in *t*+1.

The update step applied in SARSA is defined as in (1) [6]. One can observe that, besides the components listed above, it utilizes a learning factor α, that determines how much the new data influences the current estimate, and a discount factor γ to discount influence from predicted state on the current estimate.

$$Q(S_t, A_t) \leftarrow (1-\alpha)Q(S_t, A_t) + \\ +\alpha(R_{t+1} + \gamma Q(S_{t+1}, A_{t+1})) \quad (1)$$

As shown in [6], the probability of the SARSA algorithm convergence to optimal policy of the agent is 100% if the ϵ-*greedy* policy is utilized. It means that for each environment state observation, the action taken by the agent is random with a probability of ϵ, otherwise it results from the agent's policy. This approach has been utilized in the proposed algorithm.

In the context of SARSA application for traffic steering in HetNets, the environment is constituted by the considered radio access network area together with its active users. Next, the observation of the environment is a set of decision criteria for a single user. The reward after performing an action by the agent is the user's satisfaction after being served. In our approach, the ANN works as the agent, and its policy is modified according to the reward received from the environment after each action taken. This means that the ANN's connection weights and biases are modified according to (1).

The probability of a random action ϵ resulting from the adopted policy of the ϵ-*greedy* agent has been implemented in such a way that with successive simulation episodes it decreases by a certain ϵ_{dec} value, descending to zero. Such a mechanism is aimed at reducing the impact of random decisions, when the optimization of the agent's policy allows obtaining satisfactory results. Based on the trials of the RLLB method with various combinations of the SARSA method parameters, the following values of the parameters were decided: $\epsilon = 0.1$, $\epsilon_{dec} = 10^{-5}$, $\alpha = 0,15$, and $\gamma = 0,95$.

In order for the ANN to be able to optimize its behaviour an enhanced set of criteria is processed. In particular, for each user-cell pair, a set of three parameters is used: current cell load, percentage of cell's available radio resources that would be used up upon serving the user, and an estimated remaining number of users to be handled by the cell. The last parameter may be difficult to determine, however, in a real scenario, the initial process of load balancing could utilize one of the reference algorithms. Then, the number of users handled by the reference method can be used as input for the RLLB method.

Regarding the ANN structure, the selection of the number of convolution layers and the number of channels in each layer was performed on the basis of the method efficiency tests for various layer structures. For a large number of layers or channels in each layer, the efficiency of the RLLB method did not improve in relation to a network with a less complex structure. With too few layers or channels in each layer, the RLLB method lead to random decisions and its performance did not improve with subsequent simulation episodes. Finally, a decision was made on the structure of a neural network with three convolution layers with the number of channels 20, 10 and 1. Additionally, the ReLU activation function (i.e. a function that outputs the input directly if it is positive; otherwise, it outputs zero) was used in the first two layers, which significantly improved the stability of the

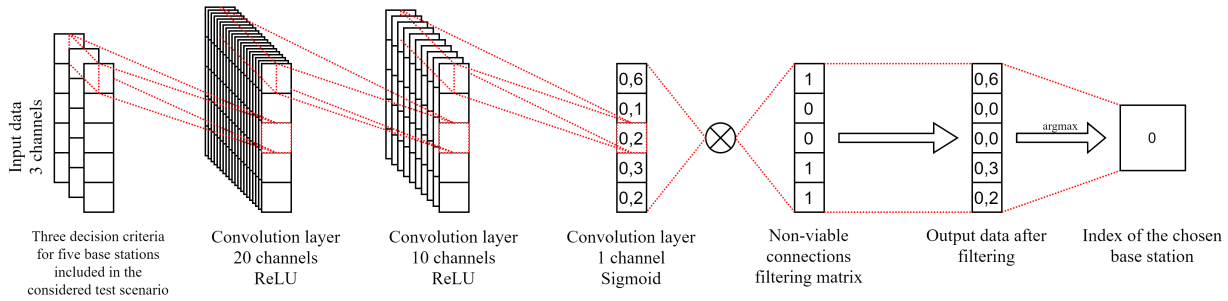

Fig. 1. Artificial neural network structure used in RLLB method

network. Next, the Sigmoid activation function (i.e. a function transforming input data to values from 0 to 1 [7]) was used in the third convolution layer, as it allows to obtain values that are convenient for interpretation. Thus, the output values of the neural network can be treated as an assessment of the goodness of the allocation of resources of a given base station to the user expressed as a percentage. Final structure of the neural network utilized in the RLLB method is shown in Figure 1.

Values at the output of the last convolution layer are subject to additional filtering to exclude base stations with insufficient signal coverage in the position of the currently considered user (i.e. channel quality indicator or user satisfaction equal to 0). The algorithm handles the user using resources of the base station with index equal to the index of the highest value after filtering the output of the ANN.

III. Scenario description

A. Base stations and user distribution

Simulation scenario for (downlink) traffic steering algorithms' efficiency comparison includes an urban environment with five base stations, including both LTE-A and NR radio access technologies. In the center, there is an LTE-A macro cell which can transmit with a maximum power of 43 dBm; two LTE-A and two NR micro cells are deployed within its coverage range, creating the HetNet scenario. The transmit power of the former cells is 32 dBm, whereas for the latter – 34 dBm. The operating frequencies for the LTE-A and NR cells are 2100 MHz and 3500 MHz, respectively.

Users are distributed randomly in range of each base station, i.e. 300 users in range of the macro cell and 60 users in range of each micro cell. Each user is assigned one of the three available user profiles (as defined in Tab. I) with specific probabilities.

TABLE I
User profile parameters in simulation scenario

Profile name	Probability	Bit rate demand
Voice (low rate)	75%	96 kbps
Data (mid rate)	20%	5 Mbps
Data (high rate)	5%	24 Mbps

B. Radio channel quality model

Standardized channel model, described in [8], was used to calculate line of sight (LOS) probabilities and pathloss for each user-cell pair. The model has been utilized for both LTE-A and NR cells, as it is declared viable for frequencies from 500 MHz to 100 GHz. As the considered simulation scenario includes an urban environment, UMa and UMi variants of the channel model are used for macro and micro cells, respectively.

Based on the calculated pathloss, transmitted power of the given cell and gains/losses related to base station's and user's equipment, a value of signal power received by the user is determined for each base station assuming the receiver noise sensitivity at -110 dBm. The SINR value is then compared with target values in the standardized CQI (e.g. [9]) to find an estimated code rate used for further effective bit rate calculations. In the simulation, 20 MHz bandwidth was considered with 15 kHz and 30 kHz of subcarrier spacing for LTE-A and NR, accordingly.

C. Traffic steering process

Each simulation episode includes an allocation of radio resources to each user. At the beginning of the episode, it is assumed that all base stations have all radio resources available, and that no user is yet served by the network. The handling order of users is random. When the resources of all

base stations in the test scenario are exhausted, the episode ends. It is assumed that the allocation of resources takes place within the network controller which has information about the load of each base station, the parameters of the signals received by the user's equipment and the bitrate demand.

IV. Simulation results

To evaluate the efficiency of the proposed solution extensive computer simulations have been carried out. Per each method 30000 episodes have been considered to guarantee statistical reliability of the results. Each method's efficiency was measured against two statistics collected at the end of each episode: mean user satisfaction (MUS), and mean not-handled-user (NHU) count. The former one ranges from 0 to 1 and is calculated as the mean ratio of demanded to received bitrate among all users. The number of NHUs is determined at the end of the episode as the number of users with a satisfaction of 0.

Efficiency statistics for CLB, SLB and RLLB method are listed in Table II, where S_{av} is the MUS after 30000 simulation episodes with σ_S being the standard deviation of mean satisfaction values and N_{av} is the mean NHU count after 30000 simulation episodes with σ_N being the standard deviation of mean NHU counts.

TABLE II
Efficiency statistics results for considered traffic steering methods in the test scenario

Method	CLB	SLB	RLLB
S_{av}	0.860	1.000	0.998
σ_S	0.020	0.000	0.002
N_{av}	74.80	75.98	73.89
σ_N	11.02	10.94	10.87

The CLB method provides the lowest MUS of 86%, with its standard deviation equal to 2%. The SLB method managed to achieve 100% MUS in every episode. The RLLB method is almost as efficient in this respect as the SLB method; it managed to achieve near 100% MUS, with its standard deviation ten times lower than in the CLB method. Next, in terms of mean NHU count, the SLB method is the least efficient, as it leaves almost 76 users not handled on average. The CLB method is capable of handling one user more on average. The most efficient method in this respect is the RLLB method, which manages to handle two users more than the SLB method on average. Moreover, the RLLB method manages to combine the best features of both reference methods, with almost 100% MUS as the SLB method and a relatively small NHU count as the CLB method.

V. Conclusion

The novel algorithm called Reinforcement Learning Load Balancing successfully uses an ANN trained with the SARSA algorithm for optimal traffic control in radio access networks. The method adjusts its decisions based on feedback from the network in form of user satisfaction observed during the simulation episode. Decisions made with use of the continuously optimized ANN model allow for achievement of a small number of NHUs and high satisfaction of the users served.

References

[1] Mamta Agiwal, Abhishek Roy, and Navrati Saxena. "Next Generation 5G Wireless Networks: A Comprehensive Survey". In: *IEEE Com. Sur. Tut.* 18.3 (2016), pp. 1617–1655.

[2] Ying Loong Lee et al. "Recent Advances in Radio Resource Management for Heterogeneous LTE/LTE-A Networks". In: *IEEE Com. Surv. & Tut., Vol. 16* 4 (2014).

[3] Dantong Liu et al. "User Association in 5G Networks: A Survey and an Outlook". In: *IEEE Com. Surv.& Tut.*. 18.2 (2016), pp. 1018–1044.

[4] Haitham Khaled et al. "A Green Traffic Steering Solution for Next Generation Communication Networks". In: *IEEE Transactions on Cognitive Communications and Networking* 7.1 (2021), pp. 222–238.

[5] Vučević Nemanja et al. "Reinforcement learning for joint radio resource management in LTE-UMTS scenarios". In: *Comp. Netw.* 55 (2011).

[6] Richard Sutton and Andrew Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2015.

[7] Simon Haykin. *Neural Networks and Learning Machines*. Pearson, 2009.

[8] 3GPP. *Study on channel model for frequencies from 0.5 to 100 GHz*. Tech. rep. Chapter 7.4, TR 38.901, V16.1.0. ETSI, 2019.

[9] *NR; Physical layer procedures for data*. TS 38.214, V16.5.0. 3GPP. 2021.