Radio Resource Allocation in 5G/B5G Networks: A Dimension Reduction Approach using Markov Decision Processes

Lucas Inglés[1][0000–0002–8125–9764], Olivier Tsemogne[2][0000–0002–0989–3269], and Claudina Rattaro[1][0000–0001–7149–5934]

[1] Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay
{lucasi,crattaro}@fing.edu.uy
[2] IMT Atlantique, Brest, France
serge-olivier.tsemogne-kamguia@imt-atlantique.fr

Abstract. We tackle radio resource allocation in 5G and B5G networks, focusing on applications with stringent delay requirements. We formulate this problem as a discounted Markov Decision Process (MDP), considering each user's Channel Quality Indicator and queue status. We introduce a reducible MDP using state abstraction. By mapping transition dynamics and rewards to an abstract state space, we simplify solving MDPs with smaller state spaces, avoiding the complexity of the original high-dimensional state space. We explore different methods for weighted state aggregation and verify through simulations that our dimension reduction strategy yields results close to the optimal policy.

Keywords: radio resource allocation · markov decision process · dimension reduction.

1 Introduction

The fifth generation (5G) of wireless networks has been purposefully designed to accommodate a wide range of network services, each with its own requirements. These 5G services are classified into three primary categories: ultra-reliable and low-latency communication (uRLLC), enhanced mobile broadband (eMBB), and massive machine-type communication (mMTC). These use cases frequently have conflicting demands, necessitating a radio design that is highly versatile and adaptable to address the varying conditions of each service category efficiently [2]. In addition to the strategic configuration of the radio interface (e.g., numerology settings), ensuring specific quality of service levels, particularly to meet stringent maximum delay requirements, heavily relies on the effectiveness of resource allocation algorithms.

In this work, we study a 5G/B5G downlink scheduling system. We model the resource allocation problem as a Markov Decision Process (i.e. the ground MDP), incorporating both the Channel Quality Indicator (CQI) of each user in each Physical Resource Block (PRB) and the queue status of each user. Our goal

is to find a scheduling policy that minimizes the queuing delay experienced by users, effectively reducing the sum-delay. Diverse applications, including real-time video streaming, online gaming, and smart transportation, underscore the critical importance of optimizing downlink scheduling and resource allocation to ensure seamless connectivity and high performance across various sectors.

Interestingly, the vast majority of works that study the resource allocation problem by modeling it as a MDP then solve this complex problem, which involves high-dimensional states and action spaces, using Artificial Intelligence techniques. In particular, Reinforcement Learning (RL) is a prevalent tool employed in these studies (see for example [6, 1, 4, 5]). Inspired by state abstraction techniques, which have been shown to significantly improve the efficiency of MDP-solving algorithms [3, 7], we formulate an abstract MDP. Our approach defines the similarity between states and works with a reduced complexity system by grouping similar states into aggregate classes. The aggregation defines transition and reward dynamics between classes.

The main contribution of this work is the introduction of an approach for solving high-dimensional Markov Decision Processes (MDPs), specifically in the field of mobile communications. We propose different abstractions to solve the original resource allocation problem efficiently. By conducting a thorough comparative analysis of different weight distribution strategies for state aggregation, we identify the most effective approximate solution for the base MDP. This analysis considers factors such as convergence time, proximity to the optimal solution (the solution of the ground MDP), and other relevant metrics. Our results show significant improvements in the resolution times of the resulting MDP, while maintaining minimal error and ensuring the extrapolation to the original model. Moreover, we provide access to our repository housing the simulations conducted, facilitating further exploration and validation of our results (GitHub Repository[3]).

The remainder of the article is structured as follows. In Section 2 we introduce our hypotheses and the main characteristics of the considered resource allocation problem. We also formulate our ground MDP and provide some intuition that will explain the state aggregation, which will be the basis of Section 3. In Section 3 we describe the different abstractions and we present some results. We conclude in Section 4.

2 Model Description

The scheduler efficiently allocates bandwidth among slices and users, using the Physical Resource Block (PRB) as its basic unit. In 5G's OFDM system, a PRB consists of 12 OFDM subcarriers and one Transmission Time Interval (TTI). In this work, we aim to reduce the state-action space involved in the scheduler's decision-making process. To start, let us first introduce the problem description.

[3] https://github.com/Tsemogne/Radio-Resource-Allocation

2.1 Problem Description

The time is divided into discrete time slots (i.e. TTIs). Our system comprises N User Equipments (UEs), denoted as $\text{UE}_1, \ldots, \text{UE}_i, \ldots, \text{UE}_N$, and M PRBs, denoted as $\text{PRB}_1, \ldots, \text{PRB}_j, \ldots, \text{PRB}_M$. We assume that the Channel Quality Indicator (CQI) for each UE in each PRB, represented as $h_{i,j}$, remains constant throughout the time. These CQI values are organized into an $N \times M$ matrix, denoted as $\mathbf{h}$, which we refer to as the Channel Quality (CQ) matrix. At each time slot, our scheduler allocates each PRB to exactly one UE for that time slot duration. This allocation can be represented as a tuple $\mathbf{a}$, where $\mathbf{a}(j)$ represents the UE to which PRB_j is allocated. In matrix form, $\mathbf{a}$ is an $M \times N$ matrix. Both matrix are represented in Eq.(1). Here, $a_{j,i}$ equals 1 if the j-th PRB is allocated to the i-th UE and 0 otherwise. Notably, each row in matrix $\mathbf{a}$ has only one non-zero entry, indicating the UE to which the corresponding PRB is allocated.

$$\mathbf{h} = \begin{bmatrix} h_{1,1} & \ldots & h_{1,j} & \ldots & h_{1,M} \\ \vdots & & \vdots & & \vdots \\ h_{i,1} & \ldots & h_{i,j} & \ldots & h_{i,M} \\ \vdots & & \vdots & & \vdots \\ h_{N,1} & \ldots & h_{N,j} & \ldots & h_{N,M} \end{bmatrix} \qquad \mathbf{a} = \begin{bmatrix} a_{1,1} & \ldots & a_{1,i} & \ldots & a_{1,N} \\ \vdots & & \vdots & & \vdots \\ a_{j,1} & \ldots & a_{j,i} & \ldots & a_{j,N} \\ \vdots & & \vdots & & \vdots \\ a_{M,1} & \ldots & a_{M,i} & \ldots & a_{M,N} \end{bmatrix} \tag{1}$$

We assume that once an PRB is allocated to a UE, it enables the transmission of $qh_{i,j}$ bits, where q is a positive constant real number. The total number of bits scheduled for transmission by the i-th UE is given by T_i:

$$T_i = q \sum_{j=1}^{M} h_{i,j} a_{j,i}. \tag{2}$$

Then, the size of the data remaining in the buffer of the i-th UE after transmission is calculated as:

$$\text{REST}_i = \max(0, b_i - T_i). \tag{3}$$

Here, b_i represents the size of data in the buffer at the beginning of the time slot[4]. After the transmission, the buffer of each UE with a maximum size of B bits receives a random number l_i of bits, following a known probability distribution. The buffer can't store more than $B - \text{REST}_i$ bits and will drop any extra bits. Therefore, the size of the data in the buffer after the time slot is:

$$b_i' = \min\left(B, l_i + \text{REST}_i\right). \tag{4}$$

The decision maker incurs two costs: one for the dropped data and another for the delay associated with the non-transmitted data:

$$c_i' = \alpha\Big(\max(0, l_i + \text{REST}_i - B)\Big)^x + \beta\left(\text{REST}_i\right)^y, \tag{5}$$

[4] We omit the temporal reference to clarify the notation. Note that REST_i represents $b_i(t+1)$, then Eq.(3) can be re-written as $b_i(t+1) = \max(0, b_i(t) - T_i(t))$..

where α, β, x, and y are positive coefficients. The first term represents the cost due to excess data that exceeds the buffer capacity, while the second term accounts for the penalty related to the delay of the data remaining in the buffer.

2.2 The ground MDP Model

Our problem can be modeled as an MDP $\mathcal{M} = (\mathcal{B}, \mathcal{A}, \mathbf{P}, \mathbf{c})$ where:

- The state of the network is the size $b = (b_i)_{i=1}^{N}$ of the data in the buffers with $0 \leqslant b_i \leqslant B$, then $\mathcal{B} = \{0, \ldots, B\}^N$;
- An action is any matrix $\mathbf{a} \in \{0,1\}^{M \times N}$ with exactly one non-zero entry in each row;
- The transition probabilities are given by

$$P(b' \big| b, \mathbf{a}) = \prod_{i=1}^{N} \mathbb{P}\left\{ l_i = b'_i - \textsc{rest}_i \right\} = \prod_{i=1}^{N} \mathbb{E}^{l} \left[\mathbb{1}_{\{b'_i = \textsc{rest}_i + l_i\}} \right] \tag{6}$$

  where $\textsc{rest}_i$ is defined in Eq. (3);
- The expected cost associated with the transition of each UE buffer is related to Eq. (5). That is, the cost function is defined by

$$c(b, \mathbf{a}) = \sum_{i=1}^{N} \mathbb{E}\Big[\alpha \Big(\max(0, l_i + \textsc{rest}_i - B) \Big)^x + \beta \, (\textsc{rest}_i)^y \Big]. \tag{7}$$

We now introduce the value function $V^{\pi}(b)$, that indicates how beneficial (or detrimental) it is to be in each state while adhering to the policy π, as follows,

$$V^{\pi}(b) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t c^t \left(b^t, \pi(b^t) \right) \middle| b = b^0 \right]. \tag{8}$$

where γ is the discount factor. Then, the scheduling problem's objective is to determine the scheduling policy that solves the optimization $\min_{\pi \in \Pi} V^{\pi}(b)$ where Π denotes the set of all possible policies. If there are a finite number of states, then in principle dynamic programming techniques obtain the optimal policy. However, the difficulty of this dynamic programming increases exponentially in the number of states, which in this case increases exponentially in the number of UEs.

2.3 Intuitions for Dimension Reduction

We can observe that the total number of possible states can be expressed as $|\mathcal{B}| = (B+1)^N$. For instance, if we have $B = 2$ and $N = 3$, the total number of states is $|\mathcal{B}| = 27$. However, if we consider an increase to $B' = 3$ and $N' = 5$, the total number of states becomes $|\mathcal{B}'| = 1024$. This substantial growth in the number of states with small increments in B and N underscores the computational

complexity of the problem. When extending this model to a real-world scenario with large values of N and B, such as in 5G network contexts, the defined model becomes unmanageable.

In order to address the exponential growing of the state space, let us first explore the dynamics of our system. In doing so, we can find that, on average, the scheduler allocates $\frac{M}{N} q H_i$ bits for transmission from the buffer of a UE with b_i bits. Therefore, the remaining bits in buffer are $\max\left(b_i - \frac{M}{N} q H_i, 0\right)$ where $H_i = \sum_{j=1}^{M} h_{i,j}$ is the average CQI of the UE. This means that, in expectation, the buffer is overloaded if $\mathbb{E}[l_i] + \max\left(b_i - \frac{M}{N} q H_i, 0\right) \geqslant B$, i.e., if

$$\mathbb{E}[l_i] > B \qquad \text{or} \qquad \begin{cases} \mathbb{E}[l_i] \leqslant B \\ b_i \leqslant B + \frac{M}{N} q H_i - \mathbb{E}[l_i] \end{cases}.$$

That is, assuming that each buffer satisfy the minimum requirement $\mathbb{E}[l_i] \leqslant B$, we can characterize an UE by the expected difference between the transmission and the arrivals as:

$$\chi(i) = \frac{M}{N} q H_i - \mathbb{E}[l_i] \tag{9}$$

In order to show the impact of the selected characteristic on the resource allocation algorithm, we have constructed a simple scenario involving three users $N = 3$, each with a maximum buffer size of 2 bits $B = 2$, and two physical resource blocks available for allocation $M = 2$. We have solved the MDP (defined in previous subsection) using the value iteration algorithm and obtained the results shown in Figure 1.

In this proposed scenario, we have assigned similar arrival rates to UE_1 and UE_2, whilst a different one to UE_0. Consequently, the characteristics of UE_1 and UE_2 are very similar. Analyzing the figure, we observe a correlation in resource allocation between UE_1 and UE_2. For the same states, UE_1 and UE_2 are more likely to receive the same amount of resources compared to UE_0 - UE_2 or UE_0 - UE_1. More precisely, UE_1 and UE_2 share the same allocation within thirteen states, while UE_0 shares eight with UE_1 and only six with UE_2. Additionally, between UE_1 and UE_2, when one is favored in the current allocation, the next allocation tends to favor the other, resulting in assignments close to the average. This artificial scenario depicts how users with similar characteristics tend to obtain similar assignations.

Considering the aforementioned, we can group the states to address the issue of the state space. We will group the states using abstractions inspired by the concepts presented here. This way, we will obtain a new MDP which consists of a whole new set of states with lower cardinality. Therefore, it becomes easier to search for a solution to the original problem in this new MDP, as it significantly reduces the computational cost. Afterward, we may finally infer the solution into the original MDP.

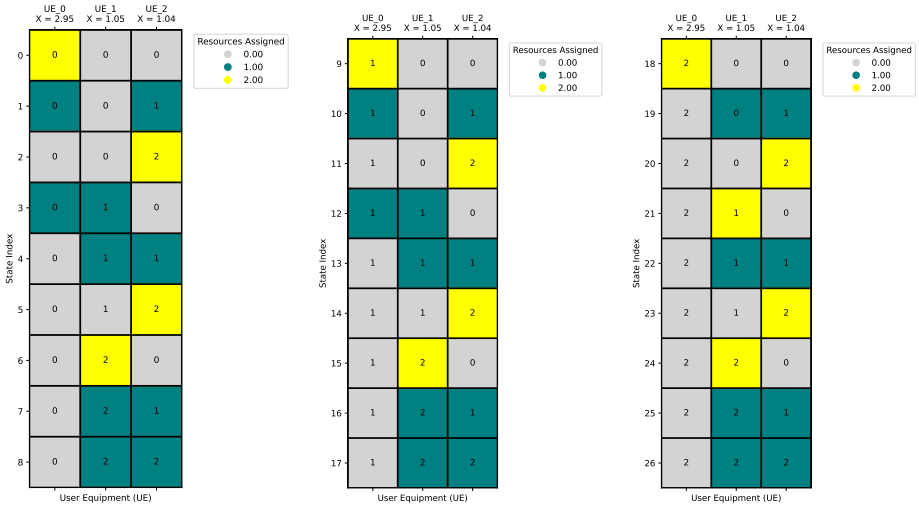

Fig. 1: Comparison of Resource Allocation Grids. The horizontal axis represents the UEs with their associated characteristics, while the vertical axis corresponds to the state index. The number in each grid element indicates the number of bits in the user's buffer and the color represents the number of resource blocks assigned to each UE for the respective state.

3 Dimension Reduction

To solve our MDP we formulate an abstract one by transfer of its dynamics on a smaller set of "abstract states" (or mega states) that correspond to classes of ground states. In this section we present the proposed abstractions and the obtained results.

3.1 Grouping the States

Motivated by the intuition presented in the previous section, a natural criterion for this grouping is to use the definition given in Eq.(9) as a characteristic, among other possible criteria. Then, all UE characteristics lay in the interval $\mathcal{K} = \left[\min_{i=1,\dots,N} \chi(i), \max_{i=1,\dots,N} \chi(i)\right]$. To group the UEs, we divide the range $\mathcal{K}$ of characteristics in a certain number K of contiguous intervals $\mathcal{K}_1, \dots, \mathcal{K}_k, \dots, \mathcal{K}_K$ by the mean of bounds $\min_{i=1,\dots,N} \chi(i) = \beta_0 < \dots < \beta_k < \dots < \beta_K = \max_{i=1,\dots,N} \chi(i)$, by posing $\begin{cases} \mathcal{K}_k = [\beta_{k-1}, \beta_k] & \text{if } k < K \\ \mathcal{K}_K = [\beta_{K-1}, \beta_K] \end{cases}$. Now we group UEs of which characteristics lay in the same division. So, the k-th group of UEs is $\text{GUE}_k = \chi^{-1}(\mathcal{K}_k) = \{i = 1,\dots,N \,|\, \beta_{k-1} \leqslant \chi(i) < \beta_k\}$ if $k < K$, or $\text{GUE}_k = \{i = 1,\dots,N \,|\, \beta_{K-1} \leqslant \chi(i) \leqslant \beta_K\}$ if $k = K$. Finally, we group the states according to the total number $\phi_k(b) =$

$\sum_{i\in \text{GUE}_k} b_i$ of bits in each group of UE's buffer. The class of a state b is thereby determined by the K-tuple $\phi(b) = (\phi_1(b), \ldots, \phi_k(b), \ldots, \phi_K(b))$. Clearly, ϕ is an abstraction that takes values in the set $\mathcal{U} = \bigtimes_{k=1}^{K} \{0, \ldots, B_k\}$, where $B_k = |\text{GUE}_k|\, B$, and $\mathcal{U}$ is henceforth the abstract state space. The number of abstract states is $|\mathcal{U}| = \prod_{k=1}^{K} (1 + |\text{GUE}_k|\, B)$, the maximum value being achieved when the numbers of UEs in two UE groups differ from at most 1.

3.2 Approximated Solution

Assuming that a weight distribution $\omega\colon \mathcal{B} \longrightarrow \mathbb{R}_+$ is set on the classes of states, i.e., $\sum_{b\in\phi^{-1}} \omega(b)$, the transition and the cost dynamics are transferred on $\mathcal{U}$ as: $\overline{P}(u'|u,a) = \sum_{b\in\phi^{-1}(u)} \omega(b) \sum_{b'\in\phi^{-1}(u')} P(b'|b,a)$ and $\overline{c}(u,a) = \sum_{b\in\phi^{-1}(u)} \omega(b)\, c(b,a)$. Each policy μ of the so defined MDP $\left(\mathcal{U}, \mathcal{A}, \overline{P}, \overline{c}\right)$ induces the policy π of the ground MDP $(\mathcal{B}, \mathcal{A}, P, c)$ defined by constant extrapolation, i.e., as $\pi(s) = \mu(\phi(s))$. If μ is the optimal policy of $\left(\mathcal{U}, \mathcal{A}, \overline{P}, \overline{c}\right)$, then π is a quasi-optimal policy of $(\mathcal{B}, \mathcal{A}, P, c)$.

3.3 Weight Distribution in Classes

Among many possibilities, we randomly select a representative state in each class or we weight the states according to the similarity (or dissimilarity) in their components. The idea behind this is to weight according to the extent to which the groups of UEs are homogeneous. We examine the impact of the following weight distributions:

Weighting the States after the UEs . We associate an N-tuple $(\eta_i)_{i=1}^{N}$ of coefficients (that need not sum to 1) with the UEs. This tuple induces a coefficient $\text{COEF}^{[\eta]}(b)$ and a weight $\omega^{[\eta]}(b)$ for each state of the MDP, defined as $\text{COEF}^{[\eta]}(b) = \sum_{i=1}^{N} \eta_i b_i$ and $\omega^{[\eta]}(b) = \frac{\text{COEF}^{[\eta]}(b)}{\sum_{b'\in\phi^{-1}(\phi(b))} \text{COEF}^{[\eta]}(b')}$. We chose the coefficients η_i in order to capture the similarity or the dissimilarity of the sizes of queues of the same group at each time slot.

- To capture the similarity, we take the size, $\eta_i = b_i$, or the closeness, $\eta_i = e^{\left|b_i - \frac{1}{|\text{GUE}_k|}\sum_{j\in\text{GUE}_k} b_j\right|}$ between the UE and the average of its group. We call these models respectively the UE-based empirical and the UE-based closeness models.
- To capture the dissimilarity, we take the distance $\eta_i = \left|b_i - \frac{1}{|\text{GUE}_k|}\sum_{j\in\text{GUE}_k} b_j\right|$ between the UE and the average of its group. We call this model the UE-based distance model.

Directly Weighting the States . We associate each group of UEs with a coefficient $\text{COEF}_k(b)$ that captures the similarity or the dissimilarity of the sizes of its

members. Then, we aggregate the coefficients of each state and normalize all the aggregated coefficients to obtain a weight ω defined as $\text{COEF}(b) = \sum_{k=1}^{K} \text{COEF}_k(b)$ and $\omega(b) = \frac{\text{COEF}(b)}{\sum_{b' \in \phi^{-1}(\phi(b))} \text{COEF}(b')}$. We eventually need to normalize the values b_i to $\beta_i = \frac{b_i}{\sum_{j \in \text{GUE}_k} b_j}$ before the computation of the coefficient $\text{COEF}_k(b)$. This normalization is impossible if $b_i = 0$ in all the group. This is, only one possibility represents the group and, accordingly we assign it the coefficient $\text{COEF}_k(b) = 0$.

- To capture the similarity, among other indexes, we have the cosine similarity calculated as the basis is the equal distribution in the group, $\text{EQUAL}(b)_i = \frac{\sum_{j \in \text{GUE}_k} b_j}{|\text{GUE}_k|}$. Its value is : $\text{COEF}_k(b) = \frac{\sum_{i \in \text{GUE}_k} b_i \text{EQUAL}(b)_i}{\sqrt{\sum_{i \in \text{GUE}_k} b_i^2} \sqrt{\sum_{i \in \text{GUE}_k} \text{EQUAL}(b)_i^2}}$. We call this weighting model the (state-based) cosine similarity model.
- To capture the dissimilarity, we perform the state-based: standard deviation model with $\text{COEF}_k(b) = \frac{\sqrt{\sum_{i \in \text{GUE}_k} \left(b_i - \text{EQUAL}(b)_i\right)^2}}{\sum_{i \in \text{GUE}_k} b_i}$; cross entropy model with $\text{COEF}_k(b) = -\sum_{i \in \text{GUE}_k} \beta_i \ln\left(\text{EQUAL}(\beta)_i\right) = \ln|\text{GUE}_k|$; and total difference with the Gini index $\text{COEF}_k(b) = \frac{\sum_{i,j \in \text{GUE}_k} |b_i - b_j|}{2|\text{GUE}_k| \sum_{i \in \text{GUE}_k} b_i}$.

Representative Selection . Another weighting model consists in choosing a representative in each class, which is equivalent to assigning some member of the class the weight value 1 and no weight to the other members. We do it either randomly or on the basis of the above criteria. We name the first model random representative selection, while the other models are the criterion-based representative selection. We distinguish between the criterion-based one representative-selection that consists in randomly selection a representative that maximizes the underlined criterion, and the criterion-based all representative-selection that equally weights all the representatives that maximize the underlined criterion.

3.4 Results

We conduct several numerical evaluations to assess the performance of the proposed abstractions[5]. Different simulations can be run in our available GitHub Repository by changing the model parameters. In all cases, promising results are obtained, significantly reducing the complexity of the problem, which translates into a notable reduction in execution times. Although an error analysis of this approach is not performed, it is shown that the state abstraction works.

In particular, we work with a scenario composed of $N = 4$, $B = 3$, and $M = 2$, solving the ground MDP using the classical value iteration algorithm and obtained the precise solution and the optimal policy. Additionally, we explore

[5] All simulations were performed using an Intel Core i7, 11th Generation, 8-core, 2.8 GHz processor with 32 GB of RAM

different aggregations; Table 1 summarizes the results obtained by setting the number of groups to two. Each abstraction model (each row in table 1) is characterized by the number of states selected in each class (select_mode: one top-weighted state, the top-weighted states, or all states), the criteria used to weight the states (groups or UEs), the rule determining the state weights (uniform distribution, similarity, or dissimilarity), and the variant (standard deviation, cross entropy or Gini coefficient) of this rule when the items receiving coefficients were groups and the rule referred to dissimilarity. Then, columns six and seven indicate the maximum differences between the precise and approximated solutions, as well as the greatest divergences between the exact and approximated optimal policies. The table also presents the resolution times and abstraction times, and it indicates in the last column the percentage of the resolution time (including the extrapolation time) relative to the resolution time of the ground model.

Table 1: Comparison of abstraction models. Parameters: $N = 4$, $B = 3$, $M = 2$ and $\gamma = 0.9$. The precision for each MDP resolution was set to 10^{-16}. Cost function parameters $x = y = \alpha = \beta = 1$. % of total_resolution_time is the percentage relative to the resolution time of the ground model.

ID	coef_owners	coef_criterion	criterion_variant	select_mode	max_diff_values	max_diff_actions	abstraction_time	resolution_time	extrapolation_time	% of total_resolution_time
1	UEs	uniform	–	one	7.516	15	18.680	4.786	0.002	27.363
2	UEs	uniform	–	top	10.350	15	19.097	5.104	0.006	29.201
3	UEs	uniform	–	all	10.350	15	19.041	5.125	0.003	29.301
4	UEs	sim	–	one	10.380	10	19.292	4.826	0.002	27.592
5	UEs	sim	–	top	10.793	15	21.690	5.788	0.003	33.089
6	UEs	sim	–	all	8.871	15	21.032	5.015	0.003	28.671
7	UEs	dissim	–	one	24.059	15	20.001	5.169	0.004	29.558
8	UEs	dissim	–	top	17.482	15	21.510	6.980	0.003	39.905
9	UEs	dissim	–	all	14.713	15	30.658	7.856	0.010	44.950
10	groups	uniform	–	one	9.000	15	26.170	5.122	0.003	29.283
11	groups	uniform	–	top	10.350	15	20.800	5.414	0.003	30.956
12	groups	uniform	–	all	10.350	15	25.976	7.309	0.003	41.783
13	groups	sim	–	one	13.111	12	26.534	5.331	0.003	30.479
14	groups	sim	–	top	10.793	15	21.121	5.397	0.004	30.864
15	groups	sim	–	all	9.981	15	24.329	7.064	0.006	40.396
16	groups	dissim	sd	one	13.770	15	28.580	5.554	0.003	31.750
17	groups	dissim	cross	one	11.839	15	21.104	5.539	0.003	31.670
18	groups	dissim	gini	one	8.424	15	21.075	5.354	0.003	30.608
19	groups	dissim	sd	top	17.482	15	21.782	6.519	0.004	37.272
20	groups	dissim	cross	top	10.350	15	29.282	7.550	0.003	43.162
21	groups	dissim	gini	top	17.482	15	26.261	5.403	0.003	30.888
22	groups	dissim	sd	all	14.765	15	20.632	5.367	0.005	30.702
23	groups	dissim	cross	all	10.350	15	20.749	6.982	0.004	39.918
24	groups	dissim	gini	all	14.765	15	25.828	7.118	0.003	40.697

Within this set of possible abstractions for the given problem, the first row of the table, corresponding to an abstraction where the coefficients are associated with the UEs and a uniform distribution is used, appears to be the best option in terms of resolution time and proximity to the optimal solution. The results show that for this abstraction, the resolution time is 27.3 % of the resolution time of the original MDP. This underscores the motivation to pursue this approach and achieve its utilization for near real-time decision-making.

4 Conclusions

In this work, we studied a radio resource allocation system modeled using a Markov Decision Process (MDP). Various weighted abstractions of state spaces

in MDP were presented, and simulations were conducted to compare the performance of each model against the original. The results demonstrate the potential of these abstractions in efficiently solving complex MDPs. For future work, we plan to explore more complex scenarios and utilize our abstractions to enhance the performance of artificial intelligence algorithms. Additionally, a more detailed analysis of the approximation error is necessary. Our work in progress include the combination of state and action spaces abstraction for a faster resolution. This allows accounting the variable numerology, variable Channel Quality Indicator (CQI) over time to emulate mobile users, and other factors that reflect real-world conditions more accurately. Our goal to come out with an efficient online optimization resource allocation in 5G and B5G networks.

Acknowledgments. This work was partially funded by Universidad de la República's CSIC R&D Project "5/6G Optical Network Convergence: an holistic view", a CAP PhD scholarship and a STIC/AMSUD project between CAPES/BR (88881.694462/2022-01); Ministry for Europe and Foreign Affairs/FR; Campus France/FR and the National Agency for Research and Innovation/UY (MOV_CO_2022_9_1012442).

References

1. Boutiba, K., Bagaa, M., Ksentini, A.: Optimal radio resource management in 5g nr featuring network slicing. Computer Networks **234**, 109937 (2023). https://doi.org/https://doi.org/10.1016/j.comnet.2023.109937, https://www.sciencedirect.com/science/article/pii/S1389128623003821
2. Dahlman, E., Parkvall, S., Skold, J.: 5G NR: The Next Generation Wireless Access Technology. Academic Press, 1st edn. (August 9 2018)
3. García, J., Álvaro Visús, Fernández, F.: A taxonomy for similarity metrics between markov decision processes. Machine Learning **111**(11), 4217–4247 (Nov 2022). https://doi.org/10.1007/s10994-022-06242-4, https://doi.org/10.1007/s10994-022-06242-4
4. Gu, Z., She, C., Hardjawana, W., Lumb, S., McKechnie, D., Essery, T., Vucetic, B.: Knowledge-assisted deep reinforcement learning in 5g scheduler design: From theoretical framework to implementation. IEEE Journal on Selected Areas in Communications **39**(7), 2014–2028 (2021). https://doi.org/10.1109/JSAC.2021.3078498
5. Haque, M.E., Tariq, F., Khandaker, M.R.A., Wong, K.K., Zhang, Y.: A survey of scheduling in 5g urllc and outlook for emerging 6g systems. IEEE Access **11**, 34372–34396 (2023). https://doi.org/10.1109/ACCESS.2023.3264592
6. Sharma, N., Zhang, S., Somayajula Venkata, S.R., Malandra, F., Mastronarde, N., Chakareski, J.: Deep reinforcement learning for delay-sensitive lte downlink scheduling. In: 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications. pp. 1–6 (2020). https://doi.org/10.1109/PIMRC48278.2020.9217110
7. Subramanian, J., Sinha, A., Seraj, R., Mahajan, A.: Approximate information state for approximate planning and reinforcement learning in partially observed systems. Journal of Machine Learning Research **23**(12), 1–83 (2022), http://jmlr.org/papers/v23/20-1165.html