Online Radio Access Technology Selection Algorithms in a 5G Multi-RAT Network

Arghyadip Roy, *Member, IEEE,* Prasanna Chaporkar, *Member, IEEE,* Abhay Karandikar, *Member, IEEE,* and Pranav Jha, *Member, IEEE*

Abstract—In today's wireless networks, a variety of Radio Access Technologies (RATs) are present. However, each RAT being controlled individually leads to suboptimal utilization of network resources. Due to the remarkable growth of data traffic, interworking among different RATs is becoming necessary to overcome the problem of suboptimal resource utilization. Users can be offloaded from one RAT to another based on loads of different networks, channel conditions and priority of users. We consider the optimal RAT selection problem in a Fifth Generation (5G) New Radio (NR)-Wireless Fidelity (WiFi) network where we aim to maximize the total system throughput subject to constraints on the blocking probability of high priority users and the offloading probability of low priority users. The problem is formulated as a Constrained Markov Decision Process (CMDP). We reduce the effective dimensionality of the action space by eliminating the provably suboptimal actions. We propose low-complexity online heuristics for RAT selection which can operate without the knowledge regarding the statistics of system dynamics. Network Simulator-3 (ns-3) simulations reveal that the proposed algorithms outperform traditional RAT selection algorithms under realistic network scenarios including user mobility.

Index Terms—User association, 5G NR, CMDP, WiFi offloading

1 Introduction

In the recent years, the number of mobile subscribers has increased exponentially with a rise in the popularity of data-intensive applications such as video, social networking. To address the problem of increasing data traffic consumption and demand for high data rate, small cells are being deployed by network operators. Moreover, interworking with IEEE 802.11 based Wireless Local Area Network (WLAN) (popularly known as Wireless Fidelity (WiFi) network) Access Points (APs) is also becoming popular. The reason behind this is twofold. First, WiFi AP deployment is low cost as they operate in unlicensed band. Moreover, while the Third Generation Partnership Project (3GPP) Fourth Generation (4G) Long Term Evolution (LTE) base stations and Fifth Generation (5G) New Radio (NR) [2], [3] Next Generation NodeBs (gNBs) primarily target to provide ubiquitous coverage to support high speed mobility of users, WiFi APs aim to provide high data rate in hotspot regions, campuses and homes. Such networks where different types of Radio Access Technologies (RATs) are present and a user can be associated with any RAT, are known as Heterogeneous Networks (HetNets). Therefore, for an efficient interworking between various RATs in a 5G based HetNet, the need for a unified framework enabling a global view of different RATs where control and management decisions are taken by a common entity, becomes even more important. In the absence of a global view, the utilization of network resources may become suboptimal. The 3GPP 5G standards [2] define a centralized core network to handle multiple RATs in a unified manner. Additionally, in 3GPP 5G Release 15 [2], Non 3GPP Interworking Function (N3IWF) is standardized for seamlessly integrating non-3GPP RATs such as WLAN with the centralized 5G core. Even though 3GPP 5G Release 15 standardization introduces a unified core [2], Radio Access Network (RAN) level decisions are still taken in a RAT-specific manner. However, optimal performance can be obtained if the common RAN functionalities of different RATs such as RAT selection, user offload, admission control and mobility management are controlled and managed within a unified framework. To this end, 3GPP Release 16 standard [2] introduces an Access Traffic Steering, Switching and Splitting (ATSSS) functionality which allows traffic steering among multiple RATs. Using ATSSS, network-provided policy and RAN level information from users, the centralized 5G core is able to support RAT selection and user offload in a 5G based HetNet.

We consider a 5G NR-WiFi HetNet where users of different priorities are present. Control and management functionalities of these RATs are unified at the centralized controller (for example, a Software Defined Network [4], [5] controller) in 5G core. In this paper, the centralized controller takes control and management decisions within a unified framework. Among the RAN functionalities, we consider the RAT selection problem. We consider two classes of users. Users of delay sensitive applications, such as Voice over Internet Protocol (VoIP) users are classified as high priority users, whereas users receiving best-effort service are categorized as low priority ones. We assume that high

This paper is a substantially expanded and revised version of the work in [1].

- *Arghyadip Roy was with Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, 400076, India when the work was done and is currently with Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, 61820, USA. e-mail: arghyad4@illionois.edu. Prasanna Chaporkar, Abhay Karandikar and Pranav Jha are with the Department of Electrical Engineering, Indian Institute of Technology Bombay. e-mail:{chaporkar,karandi,pranavjha}@ee.iitb.ac.in. Abhay Karandikar is currently Director, Indian Institute of Technology Kanpur (on leave from IIT Bombay), Kanpur, 208016, India. e-mail:karandi@iitk.ac.in*

priority users are served using 5G NR since WiFi may not provide the required Quality of Service (QoS). Low priority users may be served using either WiFi or 5G NR.

In our earlier works [6], [7], [8], we have addressed the trade-off between the total system throughput and the blocking probability of high priority users, in the context of an LTE based HetNet. However, in an LTE-WiFi HetNet, implementation of centralized RAT selection algorithms may be cumbersome as it requires the existence of additional network elements (which are not standardized by 3GPP), viz., a centralized controller and interfaces between LTE Evolved Node B (eNB) /WiFi AP and the centralized controller. On the contrary, in 3GPP 5G network [2], the presence of unified core enables the integration of multiple RATs. Moreover, due to the presence of N3IWF, radio related information such as load information of RATs and channel condition of users, can be gathered with ease at the centralized controller. Motivated by this, in this paper, we consider the trade-off between the total system throughput and the blocking probability of high priority users in a 5G NR-WiFi HetNet. Since low priority users are best effort in nature, blocking probability of low priority users need not be taken into consideration. However, maximizing the total system throughput subject to a blocking probability constraint may lead to a 'ping-pong' kind of behavior since the optimal policy may result in offloading [9] of a low priority user from 5G NR to WiFi and back to 5G NR again within a short time interval when a high priority user arrival is followed by a departure. Similar instances can occur where a departure is followed by an arrival. Frequent offloading between 5G NR gNB and WiFi AP causes additional delay and hence, loss of throughput. Similar problem arises for concurrent access to multiple gNBs/APs belonging to different RATs in 5G. This is due to the controller queuing delay and controller-gNB/N3IWF (AP) communication delay if there are excessive traffic steering requests (changes in the fractions of traffic through each AP/gNB) at the controller. Therefore, to address the issue of control signaling traffic in the backhaul and additional delay due to this ping-pong behavior, we incorporate an additional constraint on the offloading probability of low priority users (i.e., fraction of offloaded low priority users). We thus aim to maximize the total system throughput subject to the high priority user blocking probability and the low priority user offloading probability constraints.

The above problem is modeled as a Constrained Markov Decision Process (CMDP) problem. We establish the sub-optimality of various actions in different states of the system and thereby, reduce the effective dimensionality of the action space. However, even after reducing the size of the action space, conventional Dynamic Programming (DP) methods to solve the CMDP problem are computationally prohibitive under large state spaces. Moreover, DP methods require the knowledge of transition probabilities between different states which depend on the unknown statistics of system dynamics, viz., the arrival rates of low and high priority users. These are hard to gather in reality. To address these issues, we propose two online RAT selection heuristic algorithms. Unlike DP based algorithms, the proposed algorithms do not require the knowledge of the statistics of system dynamics. Moreover, the proposed algorithms have low computational and storage complexities. These features make the algorithms suitable for practical online implementation.

We implement the proposed RAT selection algorithms in a Network Simulator-3 (ns-3) (a discrete event network simulator) [10] based 5G NR simulation setup [11], [12]. The setup incorporates Physical (PHY) and Medium Access Control (MAC) layers of the 5G stack. The higher layers are extensions of corresponding layers in ns-3 LTE module [13]. Performances of the proposed algorithms are compared with the traditional RAT selection scheme under various practical scenarios including user mobility.

1.1 Related Work

RAT selection and offloading are among the control plane functionalities which are traditionally implemented either in distributed [14], [15], [16], [17], [18], [19], [20], [21] or centralized [6], [7], [8], [22], [23], [24], [25], [26] manner. An overview of existing RAT selection techniques in HetNets and their performance evaluation is presented in [27].

Among centralized RAT selection strategies, an integrated interference management and user association [1] problem in a two-tier HetNet is considered in [23]. Although the authors propose a computationally efficient algorithm, this approach is not adaptable to fast changes in network parameters. In [26], an admission control algorithm which maximizes the users' quality of experience, is proposed in a macro cell-small cell HetNet. The authors show that the optimal policy performs better than the random policy. In our earlier work [6], we propose a computationally efficient network-initiated RAT selection algorithm which maximizes the total system throughput subject to a blocking probability constraint. However, it requires the knowledge of the state transition probabilities of the underlying Markov chain which depend on the statistics of unknown system dynamics. Subsequently, we propose learning algorithms in [7] which can work without the knowledge of the statistics of unknown system dynamics and unlike [6], can be implemented online. The convergence speed of the traditional Q-learning based algorithm in [7] is further improved in [8] by exploiting the structural properties of the optimal policy.

Among the distributed solutions, the authors in [14] propose an association scheme which maximizes the network utility subject to constraints on user requirements. The proposed scheme is based on the utility obtained from past associations of users. In [19], a low complexity RAT selection algorithm is proposed for an LTE network comprising macro, pico and femto cells. The proposed algorithm achieves a near-optimal performance with a theoretical guarantee on the performance. The authors in [21] use the information provided by the network to improve the efficiency of distributed RAT selection algorithms.

Contrary to distributed approaches which focus on optimizing individual user utilities and hence, often may not provide the globally optimal solution, centralized approaches provide a framework for overall system optimization. Moreover, in [28], the authors demonstrate that

1. The terminologies "association" and "RAT selection" are used interchangeably throughout this paper.

network-centric resource allocation approaches perform better than distributed approaches in a HetNet. Hence, we focus on network-initiated centralized approaches for RAT selection and offloading in a 5G NR-WiFi network. The trade-off involving the total system throughput, the blocking probability and the offloading probability in a dynamic 5G NR based HetNet within an optimization framework has not been considered in the literature before. Furthermore, unlike others, we investigate the role of offloading in improving the system performance at time instances of arrivals and departures of users.

1.2 Our Contributions

In this paper, We consider the problem of optimal RAT selection in a 5G NR-WiFi HetNet where we aim to maximize the total system throughput subject to constraints on the high priority user blocking probability and the low priority user offloading probability. Our contributions are summarized as follows:

- Optimal association problem of maximizing the total system throughput subject to constraints on the high priority user blocking probability and the low priority user offloading probability is formulated as a CMDP problem.
- We prove the sub-optimality of certain actions in different states. This reduces the size of the effective action space.
- We propose two low complexity heuristics for RAT selection. They do not require the knowledge of the user arrival rates and hence, can be implemented online.
- We implement the RAT selection algorithms in a 5G NR based simulation setup [11], [12] in ns-3.
- We also compare the performances of the proposed algorithms with that of traditional RAT selection algorithm under realistic scenarios including user mobility.

The trade-off involving total system throughput, blocking probability of high priority users and offloading probability of low priority users in a dynamic 5G NR-WiFi network (where users arrive and depart) within an optimization framework is not considered in the literature before. Moreover, we investigate the role of offloading coupled with RAT selection at the arrival and departure instants of users. While the trade-off between the total system throughput and the blocking probability is addressed in [6], [7], [8], this is the first work where the constraint on the offloading probability is considered to control the frequency of switching between RATs and resulting issue of enhanced signaling and delay. We propose low complexity RAT selection algorithms which are free from the curses of dimensionality and modeling. Moreover, unlike learning based methods, these algorithms do not suffer from slow convergence issues, making them suitable for practical implementation.

The rest of the paper is organized as follows. Section 2 describes the system model. In Section 3, the problem formulation within the framework of CMDP is described. In Section 4, we derive the suboptimal actions and eliminate them from the action space. We describe the proposed algorithms in Section 5 with a comparison of storage and computational complexities. Performance analysis of the proposed algorithms in ns-3 is provided in Section 6. Section 7 provides key insights of the paper and concludes the paper.

2 System Model

We begin by first describing the 5G NR-WiFi network architecture.

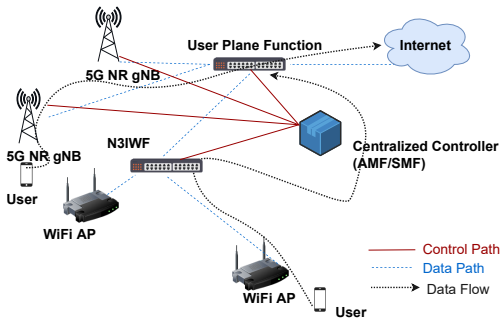


Figure 1: 5G NR-WiFi network architecture.

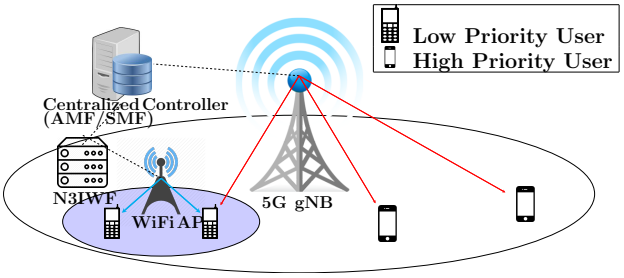


Figure 2: 5G NR-WiFi heterogeneous network.

2.1 3GPP 5G NR-WiFi Network Architecture

The centralized controller is present in the core network and controls radio access functionalities of 5G NR gNB and WiFi AP, as shown in Fig. 1. The presence of N3IWF ensures seamless integration between WLAN and 5G core. User data from 5G NR gNB and WiFi AP are routed to User Plane Function (UPF), directly and via N3IWF, respectively.

3GPP 5G standards introduce the usage of SDN technology, a unified (RAT agnostic) core and the support for ATSSS functionality [2]. SDN paradigm enables the separation of control plane functions (such as Access and Mobility Management function (AMF) and Session Management Function (SMF)) from data plane functions and usage of a logically centralized control plane. The 5G NR-WiFi network architecture in Fig. 1 exhibits both these characteristics. AMF and SMF are responsible for controlling the distributed multi-RAT (5G NR and WiFi) RAN and core network data plane functions using a single unified 5G core. Due to the centralization of control functionality in core, 5G network can support centralized RAT selection algorithms. ATSSS functionality can be used by AMF and SMF in 5G core to dynamically select one of the RATs to deliver data to the user. The RAT selection algorithm may utilize parameters such as user priority and RAT specific information, e.g., radio link quality. The RAT specific information may be collected by the controller either directly from the user or from the RAN nodes. RAT selection decision taken by the controller can be conveyed to UPF for real-time traffic distribution. Therefore, the ATSSS feature coupled with the

unified 5G core enables RAT selection and user offload from one RAT to another based on RAT specific information and user priority.

The system setting considered by us for the problem formulation consists of a 5G NR gNB and a WiFi AP. As described in Fig. 2, both 5G NR gNB and WiFi AP which is present inside the coverage area of 5G NR gNB, are connected to the centralized controller via high capacity lossless links and N3IWF.

2.2 Assumptions

The assumptions made in the system model are as follows.

High and low priority users are assumed to be present anywhere within the coverage area of 5G NR gNB. Since low priority users present outside the coverage area of WiFi AP always associates with 5G NR gNB and no decision is involved in this case, in the state space, we take into consideration only those low priority users which are present in the common coverage area of 5G NR gNB and WiFi AP. In this case, the controller needs to choose whether the user is to be associated with WiFi AP or 5G NR gNB, based on the system state.

Both high and low priority users are alloted resources in 5G NR gNB from a common set of resources. High priority users are provided a Guaranteed Bit Rate (GBR) following which available resources in 5G NR gNB are uniformly allocated among low priority users.

We assume that high priority users are always served using 5G NR since WiFi (although may provide a good channel to the user) may not provide the required QoS guarantee.

We further assume that in 5G NR, high and low priority users can be in either "good" or "bad" channel state. Based on the location of users, the coverage area of an 5G NR gNB is assumed to be divided into two regions, viz., cell center and cell edge regions [29], [30], [31]. Since cell edge users are present in the vicinity of the cell boundary or in coverage holes, usually they receive weaker signal strength than that of cell center users. Therefore, it is assumed that users present in the cell center region (outside coverage holes) have good channels, whereas cell edge users have bad channels. Cell edge users can be mapped to the 5^{th} percentile users [32], [33] for which the spectral efficiency is 5% point of the cumulative distribution function of normalized user throughput. Rest of the users can be referred to as cell center users. Although the coverage of a 5G NR gNB may be small (especially for mmwave cells), the variation in SNR may be large [34] due to factors such as blockage, reflector, small variation in handset orientation relative to the environment.

Cell center/ cell edge region can be chosen based on the average Channel Quality Indicator (CQI) experienced by the users in 5G NR, similar to [34]. If the average CQI of a user exceeds a certain threshold, then the user is called a cell center user. Otherwise, it is called a cell edge user. Although 3GPP [35] standardizes 16 possible CQI values, for analytical tractability we assume that possible CQI values can be categorized in two groups.

Since RAT selection decisions are made for a sufficiently long period of time, we assume that users are distributed in cell edge/cell center region depending on their average radio conditions. We assume that instantaneous fading effects are averaged out over the timescale in which decisions are taken.

We consider that the users are stationary, and the channel state of a user does not change with time once the user is admitted. The channel states of incoming users are assumed to be known at the centralized controller, however, the channel states in 5G NR gNB are random (good/bad) with finite probabilities.

Since WiFi AP has a small coverage area, it is assumed that the channel states of users in WiFi are always good.

Both 5G NR gNB and WiFi AP forward the channel condition of individual users to the centralized controller which takes the RAT selection decisions. If WiFi AP does not forward channel condition of a given user, the centralized controller assumes that the user cannot be associated with WiFi AP.

Let high and low priority user arrivals be independent Poisson processes with means $_H$ and $_L$, respectively. Following [36], the service times for high and low priority users are assumed to be exponentially distributed with means $\frac{1}{H}$ and $\frac{1}{L}$, respectively.

We also assume that low priority users use applications such as video where the duration of a session does not depend on the number of users.

Remark 1. *One 5G NR gNB-one WiFi AP scenario (considered for brevity for notation) in this paper can be extended easily for multiple gNBs-multiple APs case. If coverage areas of gNBs (APs) do not overlap, users present inside the coverage area of each gNB (AP) can be considered as a tuple in the state space. In the case of overlapping coverage areas, the problem can be cast into the non-overlapping coverage area case, and analysis follows in a similar way. This can be performed using some simple criterion, say, mapping every geographical point to the gNB (AP) which provides the highest Signal-to-Noise Ratio (SNR). The set where multiple gNBs/APs have equal SNR is non-generic since in the corresponding parameter space, the Lebesgue measure is 0. In case of multiple gNBs, the interference management and the power control mechanisms would be present at gNBs and handsets which employ advanced coordinated communications [37], [38]. Such mechanisms are currently under standardization by 3GPP. Interference constraints associated with the problem, if any, can be handled at individual gNBs. Hence, these constraints need not be taken into account while optimizing the system throughput subject to constraints on blocking and offloading probabilities.*

Remark 2. *As specified in 3GPP standards [39] and International Telecommunication Union (ITU) International Mobile Telecommunications-2020 (IMT-2020) requirements [33], inter-gNB distances in case of dense urban, urban macro and rural deployments are 200, 500 and 1732 m, respectively, which are typically larger than the coverage of a WiFi AP.*

Remark 3. *Since 3GPP 5G architecture integrates different RATs using a unified core, centralized algorithms can be easily implemented. Note that even if 5G NR gNB and WiFi AP are deployed by different operators, 3GPP 5G architecture [2] guarantees that the interworking between them is supported at the core network using standardized interfaces and N3IWF.*

2.3 State Space

The system can be viewed as a controlled continuous time stochastic process $\{X(t)\}_{t \geq 0}$, similar to [6], [7], [8]. Any state s in the state space S is expressed as $s = (i_G, i_B, j_G, j_B, k_G, k_B)$, where i_G, i_B denote the number of high priority users associated with 5G NR gNB with good and bad channels in 5G NR gNB, j_G, j_B denote the number of low priority users associated with 5G NR gNB with good and bad channels in 5G NR, and k_G, k_B denote the number of low priority users associated with WiFi AP, however, with respect to 5G NR they have good and bad channels, respectively. Channel states of users in WiFi are not explicitly mentioned since channel states of users in WiFi are always assumed to be good. The arrival and departure of high and low priority users with good and bad channel states in 5G NR are taken as decision epochs. It is evident that the system changes state only at the decision epochs. Moreover, since the system is Markovian, it is sufficient to observe the system state at these decision epochs and not at other time points.

An arrival or a departure of a user with good/bad channel state in 5G NR is referred to as an event. Whenever an event happens, the system changes state. Let the set of all events be denoted by E. Let us denote the arrival of a high and low priority user with good (bad) channel by $E_1(E_3)$ and $E_2(E_4)$, respectively. Let the departure of a high and low priority user with good (bad) channel be denoted by $E_5(E_6)$ and $E_7(E_8)$, respectively. We assume that the departure of a low priority user from WiFi with good and bad channel in 5G NR are denoted by E_9 and E_{10}, respectively. Note that, the channel states of users in WiFi do not appear in the event space because the channel states of users in WiFi are assumed to be good. At every decision epoch, a decision is chosen by the controller based on the event and the current system state. Based on the decision chosen, the system moves to a different state with a finite probability.

Let the 5G NR system be composed of C_N resource blocks. We assume that $s = (i_G, i_B, j_G, j_B, k_G, k_B) \in S$ if $(i_G + p_c i_B) \leq C_N$, $(j_G + j_B) \leq N$ and $(k_G + k_B) \leq W$, where N is a sufficiently large positive integer (incorporated for analytical tractability). The first condition signifies that a high priority user is admitted only when sufficient resources are available. The first condition is under the assumption that a high priority user with bad channel requires $p_c (> 1)$ times as many resource blocks as required by a high priority user with good channel. The quantity W signifies the maximum number of users that can be supported in WiFi to guarantee a specified minimum per-user throughput. Note that the per-user throughput of WiFi decreases monotonically with load [40]. Let the GBR required by a high priority user be denoted by $R_{L,H}$. A fixed number of resource blocks are allocated to a high priority user based on the channel condition of the user. Low priority users are best-effort in nature. Therefore, the available resources in 5G NR are allocated uniformly among low priority users. The data rate obtained by a low priority user depends on the channel states of the users and the number of high priority users. We assume that the bit rate of a low priority user with bad channel is $\frac{1}{d}(d > 1)$ times that of a low priority user with good channel, where d is a constant.

2.4 Action Space

Let us denote the action space by A. Action A_1 blocks an arriving user or does nothing during a departure. Actions A_2 and A_3 correspond to association with 5G NR and WiFi, respectively. Note that actions A_1, A_2, A_3 are similar to those of [6], [7], [8]. Under action A_4, a high priority user is associated with 5G NR and a low priority user with bad channel is offloaded to WiFi. Action A_5 performs offloading of a low priority user with bad (good) channel from 5G NR (WiFi) to WiFi (5G NR) when a user departs from WiFi (5G NR). Action A_6 associates a high priority user with 5G NR and offloads a low priority user with good channel to WiFi. Action A_7 offloads a low priority user with good (bad) channel from 5G NR (WiFi) to WiFi (5G NR) when a user departs from WiFi (5G NR). In case of high and low priority user arrivals, the feasible action sets are $\{A_1, A_2, A_4, A_6\}$ and $\{A_2, A_3\}$, respectively. The feasible action set for departures comprises A_1, A_5 and A_7. Note that blocking is a feasible action for high priority users only when the system is non-empty. Low priority users are blocked only when $(j_G + j_B)$ becomes N.

Table 1: Transition Probability Table.

$a \mid E_l$	$(i'_G, i'_B, j'_G, j'_B, k'_G, k'_B)$
$A_1 \mid E \setminus (E_2 \cup E_4)^C$	$(i_G, i_B, j_G, j_B, k_G, k_B)$
$A_2 \mid E_1$	$(i_G + 1, i_B, j_G, j_B, k_G, k_B)$
$A_2 \mid E_2$	$(i_G, i_B, j_G + 1, j_B, k_G, k_B)$
$A_2 \mid E_3$	$(i_G, i_B + 1, j_G, j_B, k_G, k_B)$
$A_2 \mid E_4$	$(i_G, i_B, j_G, j_B + 1, k_G, k_B)$
$A_3 \mid E_2$	$(i_G, i_B, j_G, j_B, k_G + 1, k_B)$
$A_3 \mid E_4$	$(i_G, i_B, j_G, j_B, k_G, k_B + 1)$
$A_4 \mid E_1$	$(i_G + 1, i_B, j_G, j_B \quad 1, k_G, k_B + 1)$
$A_4 \mid E_3$	$(i_G, i_B + 1, j_G, j_B \quad 1, k_G, k_B + 1)$
$A_5 \mid (E_5 \cup \cdots \cup E_8)$	$(i_G, i_B, j_G + 1, j_B, k_G \quad 1, k_B)$
$A_5 \mid (E_9 \cup E_{10})$	$(i_G, i_B, j_G, j_B \quad 1, k_G, k_B + 1)$
$A_6 \mid E_1$	$(i_G + 1, i_B, j_G \quad 1, j_B, k_G + 1, k_B)$
$A_6 \mid E_3$	$(i_G, i_B + 1, j_G \quad 1, j_B, k_G + 1, k_B)$
$A_7 \mid (E_5 \cup \cdots \cup E_8)$	$(i_G, i_B, j_G, j_B + 1, k_G, k_B \quad 1)$
$A_7 \mid (E_9 \cup E_{10})$	$(i_G, i_B, j_G \quad 1, j_B, k_G + 1, k_B)$

Remark 4. *Although consideration of 16 CQI values standardized by 3GPP [35] mimics the practical scenario better, this complicates the system model since cardinalities of state and action spaces become larger. However, the solution methodology remains identical to the one adopted in this paper. In Section 6, we propose suitable modifications to our schemes which take into account 16 possible CQI values of users and demonstrate that the resulting schemes outperform state-of-the-art algorithms.*

2.5 Transition Probabilities

Let $\hat{s} = (i^0_G, i^0_B, j^0_G, j^0_B, k^0_G, k^0_B)$, and $\{e_{fi:1 \leq i \leq 6}\}$ be a set of 6 dimensional vectors where all elements except i^{th} element (which is '1') is zero. Let p_g denote the probability that the

channel state of the arriving user in 5G NR is good. Then,

$$p_{ss^0}(a) = \begin{cases} \frac{{}_H p_g}{v(i_G^0;i_B^0;j_G^0;j_B^0;k_G^0;k_B^0)}; & s^0 = \mathcal{S}; \\ \frac{{}_H (1-p_g)}{v(i_G^0;i_B^0;j_G^0;j_B^0;k_G^0;k_B^0)}; & s^0 = \mathcal{S}; \\ \frac{{}_L p_g}{v(i_G^0;i_B^0;j_G^0;j_B^0;k_G^0;k_B^0)}; & s^0 = \mathcal{S}; \\ \frac{{}_L (1-p_g)}{v(i_G^0;i_B^0;j_G^0;j_B^0;k_G^0;k_B^0)}; & s^0 = \mathcal{S}; \\ \frac{i_G^0 \quad {}_H}{v(i_G^0;i_B^0;j_G^0;j_B^0;k_G^0;k_B^0)}; & s^0 = \mathcal{S} \quad e_1; \\ \frac{i_B^0 \quad {}_H}{v(i_G^0;i_B^0;j_G^0;j_B^0;k_G^0;k_B^0)}; & s^0 = \mathcal{S} \quad e_2; \\ \frac{j_G^0 \quad {}_L}{v(i_G^0;i_B^0;j_G^0;j_B^0;k_G^0;k_B^0)}; & s^0 = \mathcal{S} \quad e_3; \\ \frac{j_B^0 \quad {}_L}{v(i_G^0;i_B^0;j_G^0;j_B^0;k_G^0;k_B^0)}; & s^0 = \mathcal{S} \quad e_4; \\ \frac{k_G^0 \quad {}_L}{v(i_G^0;i_B^0;j_G^0;j_B^0;k_G^0;k_B^0)}; & s^0 = \mathcal{S} \quad e_5; \\ \frac{k_B^0 \quad {}_L}{v(i_G^0;i_B^0;j_G^0;j_B^0;k_G^0;k_B^0)}; & s^0 = \mathcal{S} \quad e_6. \end{cases}$$

Values of $i_G^0; i_B^0; j_G^0; j_B^0; k_G^0; k_B^0$ as a function of action a and event E_l are described in Table 1.

2.6 Rewards and Costs

Based on the system state and the action, a finite reward rate is obtained. In WiFi, the total throughput is a function of the total load of WiFi comprising low priority users with good and bad channels in 5G NR. Let the reward rate for state s and action a be denoted by $r(s;a)$. Under full buffer traffic WiFi model [40], let $R_{W;D}(k)$ be the per-user throughput when k users are present in WiFi. $R_{W;D}(k)$ is a function of success and collision probabilities (which signify the contention-driven medium access of WiFi users) and slot times for idle, busy (due to collision) and successful transmissions. The reward rate under a state-action pair is the sum of throughput of all users in 5G NR and WiFi under an action. Let us define

$$\begin{aligned} R(i_G;i_B;j_G;j_B;k_G;k_B) &= (i_G + i_B)R_{L;H} \\ &+ \frac{(C_N - i_G - p_c i_B)}{(j_G + j_B)}(j_G + \frac{j_B}{d})R_{L;L}\mathbb{1}_{f(j_G+j_B)>0g} \\ &+ (k_G + k_B)R_{W;D}(k_G + k_B); \end{aligned} \tag{1}$$

where $R_{L;L}$ is the data rate obtained when single 5G NR resource block is allocated to a low priority data user with good channel condition. Note that the quantities p_c and d which describe the factors by which requirement of number of resource blocks grow and bit rate degrade due to bad channel, have been introduced while describing the state space. The complete description of reward rates in state s under different event-action pairs is given in Table 2. Note that the association of a high priority user with 5G NR provides QoS guarantees in terms of delay, data rate and bit error rate by allocation of dedicated bearers providing GBR.

We consider cost functions due to blocking and offloading, respectively. Let the cost rates for blocking and offloading in state s under action a be denoted by $c_b(s;a)$ and $c_o(s;a)$, respectively. Whenever a high priority user is blocked, $c_b(s;a)$ is unity, else it is zero. Therefore,

$$c_b(s;a) = \begin{cases} 1; & \text{if high priority users are blocked;} \\ 0; & \text{otherwise.} \end{cases}$$

Table 2: Reward Rate Table.

$(a\|E_l)$	$r(s;a)$
$(A_1\|\ [_{E_i 2E}\ E_i)$	$R(i_G;i_B;j_G;j_B;k_G;k_B)$
$(A_2\|E_1)$	$R(i_G+1;i_B;j_G;j_B;k_G;k_B)$
$(A_2\|E_2)$	$R(i_G;i_B+1;j_G;j_B;k_G;k_B)$
$(A_2\|E_3)$	$R(i_G;i_B;j_G+1;j_B;k_G;k_B)$
$(A_2\|E_4)$	$R(i_G;i_B;j_G;j_B+1;k_G;k_B)$
$(A_3\|E_2)$	$R(i_G;i_B;j_G;j_B;k_G+1;k_B)$
$(A_3\|E_4)$	$R(i_G;i_B;j_G;j_B;k_G;k_B+1)$
$(A_4\|E_1)$	$R(i_G+1;i_B;j_G;j_B \quad 1;k_G;k_B+1)$
$(A_4\|E_3)$	$R(i_G;i_B+1;j_G;j_B \quad 1;k_G;k_B+1)$
$(A_5\|E_5\ [:::[\ E_8)$	$R(i_G;i_B;j_G+1;j_B;k_G \quad 1;k_B)$
$(A_5\|E_9\ [\ E_{10})$	$R(i_G;i_B;j_G;j_B \quad 1;k_G;k_B+1)$
$(A_6\|E_1)$	$R(i_G+1;i_B;j_G \quad 1;j_B;k_G+1;k_B)$
$(A_6\|E_2)$	$R(i_G;i_B+1;j_G \quad 1;j_B;k_G+1;k_B)$
$(A_7\|E_5\ [:::[\ E_8)$	$R(i_G;i_B;j_G;j_B+1;k_G;k_B \quad 1)$
$(A_7\|E_9\ [\ E_{10})$	$R(i_G;i_B;j_G \quad 1;j_B;k_G+1;k_B)$

Whenever one low priority user is offloaded from one RAT to another, $c_o(s;a)$ is unity, else it is zero. Therefore,

$$c_o(s;a) = \begin{cases} 1; & \text{if } a = (A_4 jj:::jjA_7); \\ 0; & \text{otherwise.} \end{cases}$$

Remark 5. *Apart from the single user offloading as considered in the paper, one may consider offloading of multiple users from one RAT to another in the action space. This may result in an improvement in the RAT selection performance. However, offloading of multiple users from one RAT to another causes significant instantaneous control signaling in the core network. Moreover, with the consideration of multiple user offloading as feasible actions, the size of the resulting action space increases. As a result, the computational complexity of obtaining the optimal policy increases.*

3 Problem Formulation & Solution Techniques

We target to determine an association *policy* which maximizes the total system throughput subject to constraints on the blocking probability of high priority users and the offloading probability of low priority users. A policy is a mapping from a state to an action specifying the action to be chosen in a state. Arrivals and departures of high and low priority users can occur at any point in time. Therefore, the considered problem is formulated as a continuous time CMDP problem. In this case, a *stationary randomized* optimal policy, i.e., a mixture of *pure* policies with finite probabilities, is known to be optimal [41].

3.1 Problem Formulation

Let M be the set of memoryless policies. We assume that the underlying Markov chains corresponding to the memoryless policies are unichain to guarantee that the Markov chains have unique stationary distributions. Let V^M, $C^{B;M}$ and $C^{O;M}$ denote the average reward, the cost due to blocking of high priority users and the cost due to offloading of low priority users over infinite horizon under policy $M \ 2 \ M$, respectively. Let the total reward, costs due to blocking and offloading till time t be denoted by $R(t), C_B(t)$

and $C_O(t)$, respectively. The objective of the problem is as follows:

$$\begin{aligned} &\text{Maximize:} \quad V^M = \lim_{t \to \infty} \frac{1}{t}\mathbb{E}_M[R(t)]; \\ &\text{subject to:} \quad C^{B;M} = \lim_{t \to \infty} \frac{1}{t}\mathbb{E}_M[C_B(t)] \quad B_{\max} \text{ and} \\ &\qquad C^{O;M} = \lim_{t \to \infty} \frac{1}{t}\mathbb{E}_M[C_O(t)] \quad O_{\max}; \end{aligned} \tag{2}$$

where $\mathbb{E}_M$ is the expectation operator corresponding to policy M, and $B_{\max}; O_{\max}$ are constraints on the blocking probability of high priority users and the offloading probability of low priority users, respectively. As the optimal policy is stationary, the limits in Equation (2) exist.

3.2 Conversion to Discrete-Time MDP and Lagrangian Approach

Optimal policy can be obtained using a combination of Relative Value Iteration Algorithm (RVIA) [42] and Lagrangian approach [41]. The approach adopted is analogous to that of [8]. However, due to the presence of an additional constraint, in this paper, we describe the approach to capture the notational specificities. For fixed Lagrange Multipliers (LMs) $_b$ and $_o$, the equivalent unconstrained reward function is

$$r(s;a;\ _b;\ _o) = r(s;a) \quad {}_b c_b(s;a) \quad {}_o c_o(s;a):$$

DP based optimality equation for the considered Semi Markov Decision Process (SMDP) $8s; s' \ 2 \ S$ is

$$V(s) = \max_a [r(s;a;\ _b;\ _o) + \sum_{s'} p_{ss'}(a) V(s') \quad t(s;a)];$$

where $V(s); t(s;a);$ denote the value function of state $s \ 2 \ S$, the mean transition time from state s when action a is chosen and the optimal average reward of the system, respectively. Since the sojourn times are known to be exponentially distributed, this is a special case of controlled continuous time Markov chain. Therefore, the following equation holds,

$$0 = \max_a [r(s;a;\ _b;\ _o) \quad + \sum_{s'} q(s'js;a) V(s')]; \tag{3}$$

where $q(s'js;a)$ denote controlled transition rates which satisfy $q(s'js;a) \quad 0$, if $s' \neq s$ and $\sum_{s'} q(s'js;a) = 0$. We scale the transition rates by a positive quantity which makes it equivalent to time scaling. This scales the average reward for every policy, however, without changing the optimal policy. Therefore, let $q(sjs;a) \ 2 \ (0;1); 8a$ (without loss of generality). Hence, $q(s'js;a) \ 2 \ [0;1]$ if $s' \neq s$. Add $V(s)$ to both sides of Equation (3). The optimality equation for an equivalent discrete-time MDP ($fX_n g$ say) with controlled transition probabilities $p_{ss'}(a)$ is as follows:

$$V(s) = \max_a [r(s;a;\ _b;\ _o) \quad + \sum_{s'} p_{ss'}(a) V(s')]; \tag{4}$$

where $p_{ss'}(a) = q(s'js;a)$ if $s' \neq s$ and $p_{ss'}(a) = 1 + q(s'js;a)$ if $s' = s$. For the rest of the paper, we focus on the equivalent discrete-time MDP in Equation (4), instead of the original continuous-time MDP.

For fixed values of $_b$ and $_o$, we use RVIA to solve the unconstrained maximization problem (see Equation (4)) using the following scheme.

$$V_{n+1}(s) = \max_a [r(s;a;\ _b;\ _o) + \sum_{s'} p_{ss'}(a) V_n(s') \quad V_n(s\ )]; \tag{5}$$

where $V_n(s)$ is the estimate of value function of state s at n^{th} iteration, and s is a fixed state. We aim to obtain the optimal values of $_b$ and $_o$, viz., $_b$ and $_o$, which maximize the average reward subject to cost constraints. The gradient descent routines to update the values of $_b$ and $_o$ at k^{th} iteration are as follows,

$$\begin{aligned} {}_{b;k+1} &= {}_{b;k} + \frac{1}{k}(B \quad {}_{b;k} \quad B_{\max}); \\ {}_{o;k+1} &= {}_{o;k} + \frac{1}{k}(O \quad {}_{o;k} \quad O_{\max}); \end{aligned}$$

where $_{b;k}$ and $_{o;k}$ are the values of $_b$ and $_o$ at k^{th} iteration, and $B \ {}_{b;k}; O \ {}_{o;k}$ are the high priority user blocking probability and low priority user offloading probability at k^{th} iteration, respectively. Note that the optimal policy for the CMDP is a randomized policy which is randomized in at most two states [43].

4 Action Elimination

The DP equations (Equations (4) and (5)) are exploited to prove the sub-optimality of certain actions in different states. Using this, the number of actions to be considered in different states can be reduced. This fact is utilized in analyzing the computational complexities of the RAT selection algorithms, as described later. The sub-optimality of different actions is established with the help of some lemmas.

4.1 Suboptimal Actions for Departures

The subsequent lemmas describe the sub-optimality of certain actions in a subset of states. Specifically, whenever a high/low priority user departs from 5G NR, A_5 is better than A_7. Therefore, in this case, A_7 is a suboptimal action. Similarly, in case of a low priority user departure from WiFi, A_7 is a suboptimal action.

Lemma 1. *A_5 is better than A_7 in case of high/low priority user departure from 5G NR (events $E_5; E_6; E_7$ and E_8).*

Proof. Proof is given in Section 8.1. □

Lemma 2. *A_5 is better than A_7 in case of low priority user departure from WiFi (events E_9 and E_{10}).*

Proof. Proof is given in Section 8.2. □

The physical significance of Lemmas 1 and 2 is that whenever there is a departure of a user, if we choose to offload a low priority user, it is always better to choose the user with good (bad) channel condition for offloading to 5G NR (WiFi). Intuitively, since the contribution of a bad user in 5G NR towards the total system throughput is less than that of a good user in 5G NR, it is better to offload a bad user to WiFi. Since we have assumed that in WiFi, every user experiences good channel condition, the total system throughput obtained by offloading a user with bad channel

condition in 5G NR to WiFi is more than that obtained by offloading a good user. Similar argument holds for the offloading of a user with good channel condition to 5G NR.

4.2 Suboptimal Actions for Arrivals

We characterize the suboptimal action in the case of high priority user arrivals. As described in the subsequent lemma, whenever there is a high priority user arrival, then irrespective of the channel condition of the user, action A_4 is better than A_6. In other words, whenever a high priority user is associated with 5G NR, and we decide to offload an existing low priority user to WiFi, it is always better to choose a user with bad channel condition rather than choosing one with good channel.

Lemma 3. *A_4 is better than A_6 in case of high priority user arrivals (events E_1 and E_3).*

Proof. Proof is similar to Lemma 2. □

5 Proposed RAT Selection Algorithms

The CMDP problem described in Section 3 can be solved using DP techniques which are computationally prohibitive. For example, in policy iteration [42], the computational complexity (which is $O(jAj^{jSj})$) is exponential in the cardinality of the state space. This is known as the *curse of dimensionality*. Although elimination of suboptimal actions in Section 4 reduces the size of action space, still the complexity remains exponential in jSj. Furthermore, to compute the optimal policy, we need to know the state transition probabilities which are governed by the statistics of arrival processes. In practice, statistics of arrival processes may be unknown. This is known as the *curse of modeling*. Although learning based approaches [7], [8] do not require the knowledge of statistics of arrival processes, usually their convergence rates are very slow [44]. Moreover, presence of multiple constraints, as considered in our problem, usually gives rise to multiple timescales [45] where the value functions and individual LMs are updated at separate timescales. However, convergence of the iterates at the slowest timescale is too slow. In addition, the storage complexity of such learning schemes is very high. In our earlier works [7], [8], we have established that traditional learning schemes have storage complexity of at least $O(jSj)$ and slow convergence behavior.

To address these issues, we propose low-complexity algorithms for RAT selection. Unlike DP based methods, they do not require the knowledge of the statistics of arrival processes and hence, can be implemented online. Morover, they do not suffer from slow convergence and high storage complexity issues prevalent in learning based approaches.

5.1 Myopic with Constraint Satisfaction Algorithm

In this subsection, we propose an algorithm which is myopic, i.e., it optimizes based on the current reward without considering the future utility. However, the proposed Myopic with Constraint Satisfaction Algorithm (MCSA) (described in Algorithm 1) is designed in such a way that it satisfies the associated constraints on the blocking probability of high priority and the offloading probability of low priority users.

Algorithm 1 Myopic with Constraint Satisfaction Association Algorithm.

Input: $R_{L;H}$; $R_{L;L}$; $R_W(:)$; $B_{\max}$; $O_{\max}$:
1: Initialize $D \leftarrow 0, A_H \leftarrow 0, B_H \leftarrow 0, O_L \leftarrow 0, F_B \leftarrow 0, F_O \leftarrow 0$.
2: **while** TRUE **do**
3: Determine event E in the current decision epoch.
4: Set $a \leftarrow \arg\max_{a2A} r(s;a)$.
5: **if** $(E = E_2 || E_4)$ **then**
6: Select action $a = a$.
7: **else if** $(E = E_1 || E_3)$ **then**
8: $A_H \leftarrow A_H + 1$.
9: **if** $B_H > (B_{\max} - {}_B)$ **then**
10: **procedure** HP–CONSTRAINT–VIOLATION
11: If $O_L < (O_{\max} - {}_O)$ select $a = a \in \mathcal{A} \setminus A_1$.
12: Else select $a = A_2$.
13: $F_O \leftarrow I_{f a = A_4 jj A_6 g}$.
14: **end procedure**
15: **else**
16: Select action $a = A_1$.
17: **end if**
18: **procedure** UPDATE–BP–OP
19: $F_B \leftarrow I_{f a = A_1 g}$.
20: $B_H \leftarrow \frac{B_H A_H + F_B}{(A_H + 1)}$.
21: $O_L \leftarrow \frac{O_L (A_H + D) + F_O}{(A_H + D + 1)}$.
22: **end procedure**
23: **else**
24: **procedure** DEPARTURE–POLICY
25: $D \leftarrow D + 1$.
26: **if** $O_L < (O_{\max} - {}_O)$ **then**
27: Select action $a = a$.
28: **else**
29: Select action $a = A_1$.
30: **end if**
31: $F_O \leftarrow I_{f a = A_5 jj A_7 g}$.
32: $O_L \leftarrow \frac{O_L (A_H + D) + F_O}{(A_H + D + 1)}$.
33: **end procedure**
34: **end if**
35: **end while**

We first determine the event corresponding to the current decision epoch. Then we determine the best action (denoted by a) which provides the highest immediate reward (Line 4). Now, based on the event, we choose different actions. If the current event is low priority user arrival (events E_2 and E_4), then we always choose the action a , irrespective of the channel condition of the user (Line 6). Since the set of actions for low priority user arrivals (A_2 and A_3) has no direct impact on the high priority user blocking probability and the low priority user offloading probability, whenever a low priority user arrives, we always choose the action which is best in the myopic sense. However, when a high priority user arrives (events E_1 and E_3), we increment the counter which keeps track of the number of high priority user arrivals (denoted by A_H). We block the incoming high priority user if the current value of blocking probability (which is B_H) is less than $B_{\max}$ ${}_B$ (Line 16). The factor ${}_B$ is incorporated to ensure that the blocking probability of high priority users remains below $B_{\max}$ in the long run. However, if B_H exceeds $B_{\max}$ ${}_B$, then actions are chosen based on the current value of offloading probability of low priority users (denoted by O_L). If O_L is less than $O_{\max}$ ${}_O$, then the action a (Line 11) is selected. Note that, similar to the margin ${}_B$ on $B_{\max}$, we consider

a margin ϵ_O on $O_{\max}$ to guarantee that the offloading probability of low priority users is less than $O_{\max}$ in the long run. However, if the offloading probability constraint is not satisfied ($O_L > O_{\max} - \epsilon_O$), then A_2 is chosen because selection of A_4 or A_6 may increase the value of O_L (Line 12). Depending on whether action involving blocking (A_1) or offloading (A_4 and A_6) is chosen, we update the values of B_H and O_L, respectively (Line 20-21). Procedures followed in case of departures are similar. Initially, we increment the counter (denoted by D). If O_L exceeds $O_{\max} - \epsilon_O$, action A_1 is chosen since it reduces the value of O_L (Line 29). Otherwise, we act in a myopic manner (Line 27). Based on the action selected, we then update the value of O_L (Line 31-32). Since the algorithm does not need the knowledge of unknown system dynamics ϵ_H and ϵ_L, unlike DP methods, it does not suffer from the curse of modeling. Note that when the considered problem does not have a feasible solution, then except few initial iterations, actions involving blocking and offloading are never chosen.

5.2 State-aware Myopic with Constraint Satisfaction Algorithm

In this subsection, we describe the shortcomings of MCSA and propose a State-aware Myopic with Constraint Satisfaction Algorithm (SMCSA) which addresses these shortcomings.

Whenever the current values of blocking and offloading probabilities are lower than the respective constraints, the proposed MCSA blocks an incoming high priority user. Hence, when the arrival rate of high priority users is small, it may lead to unnecessary blocking of high priority users. On the other hand, the optimal policy may result in a lower value of blocking probability of high priority users than that of MCSA, depending on $B_{\max}$. In this case, the optimal policy corresponding to the unconstrained problem may result in a high priority user blocking probability which is significantly lower than $B_{\max}$. Intuitively, MCSA blindly aims to satisfy the constraints of the considered problem without a consideration of the system state. Thus, MCSA always results in high priority user blocking probability values which are close to the given constraints, irrespective of ϵ_H and ϵ_L. Due to similar reasons, MCSA results in a high value of offloading probability of low priority users which is always close to $O_{\max}$, irrespective of ϵ_H and ϵ_L.

To address this, we propose SMCSA which is described in Algorithm 2. The procedures for low priority user arrivals are same as that of Algorithm 1. In the case of high priority user arrival, when the constraints on the blocking probability and the offloading probability are not satisfied, the procedure is exactly same as that of Algorithm 1. However, when the constraints on the blocking probability and the offloading probability are met, we modify the RAT selection strategy in the following way. We divide the entire state space into multiple regions based on the number of high and low priority users in the system. Let us divide the entire state space into P regions denoted by $R_1, R_2, \cdots R_P$. For a given region $R_n (1 \leq n \leq P)$, let the probability of blocking and offloading be denoted by $q(n)$ and $p(n)$ $(0 \leq q(n) \leq 1, 0 \leq p(n) \leq 1)$, respectively, where $q(n)$ and $p(n)$ are increasing functions of n, and $q(P) = p(P) = 1$. Whenever an event happens, we determine the current state of the system and evaluate the region in which it falls. If it falls in R_n, we block (choose A_1) the user with probability $q(n)$ and accept (choose A_2) with probability $(1 - q(n))$ (Line 17). Similarly, if it falls in R_n and the optimal action involves offloading, we offload with probability $p(n)$ and choose the other action with probability $(1 - p(n))$ (Line 11-12). Similar procedures are followed for the departures. If the constraint on O_L is met and the optimal action involves offloading, we offload with probability $p(n)$ and choose the other action with probability $(1 - p(n))$ (Line 25-26). The procedures for the update of B_H and O_L are same as those of Algorithm 1.

Algorithm 2 State-aware Myopic with Constraint Satisfaction Association Algorithm.

Input: $R_{L;H}$; $R_{L;L}$; $R_W(:)$; $B_{\max}$; $O_{\max}$:

1: Initialize $A_H \leftarrow 0$, $D \leftarrow 0$, $B_H \leftarrow 0$ and $O_L \leftarrow 0$, $F_B \leftarrow 0$, $F_O \leftarrow 0$.
2: **while** TRUE **do**
3: Determine the event E in the current decision epoch and the region R_n in which the current state s falls.
4: Set $a \leftarrow \arg\max_{a \in A} r(s;a)$.
5: **if** $(E = E_2 || E_4)$ **then**
6: Choose $a = a$.
7: **else if** $(E = E_1 || E_3)$ **then**
8: $A_H \leftarrow A_H + 1$.
9: **if** $B_H > (B_{\max} - \epsilon_B)$ **then**
10: **procedure** HP–CONSTRAINT–VIOLATION–SA
11: If $O_L < (O_{\max} - \epsilon_O)$ and $a = (A_4 || A_6)$
12: Choose $a = a$ (A_2) w.p. $p(n)(1 - p(n))$.
13: Else choose $a = A_2$.
14: $F_0 \leftarrow I_{\{a = A_4 || A_6\}}$.
15: **end procedure**
16: **else**
17: Choose $a = A_1 (A_2)$ w.p. $q(n)(1 - q(n))$.
18: **end if**
19: **procedure** UPDATE–BP–OP
20: See Algorithm 1.
21: **end procedure**
22: **else**
23: **procedure** DEPARTURE–POLICY–SA
24: $D \leftarrow D + 1$.
25: **if** $O_L < (O_{\max} - \epsilon_O)$ and $a = (A_5 || A_7)$ **then**
26: Choose a (A_1) w.p. $p(n)(1 - p(n))$.
27: **else**
28: Choose action $a = A_1$.
29: **end if**
30: $F_0 \leftarrow I_{\{a = A_5 || A_7\}}$.
31: $O_L \leftarrow \frac{O_L (A_H + D) + F_O}{(A_H + D + 1)}$.
32: **end procedure**
33: **end if**
34: **end while**

The key advantage of SMCSA is that when the value of ϵ_H is low, we block the incoming high priority users with low probability. As ϵ_H increases and the system gradually fills up with high priority users, the probability of blocking increases. Hence, effectively, the system observes less blocking probability than that of MCSA, when ϵ_H is low. As ϵ_H increases, blocking probability of high priority users increases since $q(n)$ is an increasing function of n. The performance of the resulting policy in the case of SMCSA is closer to the optimal policy than that in the case of MCSA. This is because unlike MCSA, the blocking is state-dependent. The blocking probability of high priority users gradually increases with ϵ_H, similar to the optimal policy. Therefore, the problem of high blocking probability (which is close to $B_{\max}$) of high priority users for all values of

$_H$, as seen in MCSA, does not arise in SMCSA. Similar observation holds in the case of offloading probability of low priority users also. As $_L$ grows, the offloading probability of low priority users gradually rises. Similar to MCSA, SMCSA does not require the knowledge of $_H$ and $_L$ and hence, is practically implementable.

Remark 6. *Since information regrading user arrival/departure is known to the network, the load information of 5G NR gNB and WiFi AP are present in the core network and hence, does not require any extra signaling. The only additional signaling required for our schemes is due to channel state information of users. However, since channel states of stationary users do not change too frequently, the resulting signaling overhead is small.*

Remark 7. *In an LTE-WiFi HetNet, implementation of the proposed RAT selection algorithms may be difficult since it requires the presence of additional network elements which are not standardized by 3GPP. Specifically, one requires a centralized controller and new interfaces between LTE eNB (WiFi AP) and the centralized controller. However, in 3GPP 5G network [2], the presence of unified core enables the integration of multiple RATs. Moreover, radio related information such as load information of RATs and channel condition of users, can be made available to the centralized controller using N3IWF.*

Remark 8. *In this paper, we consider one 5G NR gNB-one WiFi AP scenario. However, the proposed algorithms can work in the presence of multiple gNBs (considering the dense network scenario in 5G) and APs. In case of non-overlapping coverage areas, for each user, the problem reduces to one gNB-one AP scenario and the algorithms are applicable without any modification. When the coverage areas of multiple gNBs (APs) overlap, we map every user to the gNB (AP) providing highest SNR. Thus, we can tessellate the entire geographical area into multiple non-overlapping areas. For example, if a user connects to its closest AP, then Voronoi tessellation gives non-overlapping areas for the APs, satisfying our condition. Such models are adopted in [46], [47]. Following that, MCSA and SMCSA choose appropriate RAT (5G NR gNB/WiFi AP) for a particular user.*

5.3 Extensions to Multiple User Offloading[2]

The proposed MCSA and SMCSA consider only single user offloading in the feasible action space as multiple user offloading may lead to increase in computational complexity along with an increase in instantaneous control signaling in the backhaul. However, the proposed algorithms can be extended to the case where more than one user can be offloaded. The motivation behind the consideration of multiple user offloading is when O_L is less than $O_{\max}$ $_O$, we act in a myopic fashion. In principle, offloading of multiple users may give rise to more instantaneous reward than single user offloading depending on the system state. For example, when event E_1 occurs, offloading of two low priority users with bad channels in 5G NR is better than offloading of one low priority user with bad channel in 5G NR if $R(i_G + 1; i_B; j_G; j_B \quad 1; k_G; k_B + 1) < R(i_G + 1; i_B; j_G; j_B \quad 2; k_G; k_B + 2)$ (See Equation (1)). Therefore, the system performance by offloading multiple users may be better than that of single user offloading.

2. We thank the reviewer for pointing this out.

The modifications to MCSA which are necessary to take into account offloading of multiple users, are as follows. Let the resulting action space after incorporation of multiple user offloading be denoted as A'. Now, the best action a is chosen as $a \quad \arg\max_{a2A'} r(s;a)$ and $a \quad \arg\max_{a2A' \cap A_1} r(s;a)$ (Lines 4 and 11 of Algorithm 1, respectively). Additionally, if a dictates the offload of u users (say), then we use the update rule $F_0 \quad u$ (Lines 13 and 31 of Algorithm 1). The rest of the algorithm remains unmodified. The modifications required in the case SMCSA are similar to the foregoing modifications required for MCSA.

5.4 Comparison of Complexities

In this subsection, we analyze the computational and the storage complexities associated with the proposed algorithms and the optimal policy. We summarize the complexities of MCSA and SMCSA in Table 3. Storing the optimal action for every state results in a storage complexity of $O(jSj)$. Furthermore, the computation of the optimal policy using policy iteration has the worst case computational complexity of $O(jAj^{jSj})$, making it computationally restrictive.

In MCSA, based on every event, we need to calculate the best action a , resulting in a per-iteration computational complexity of $O(jAj)$. As discussed in Section 4, action elimination reduces the effective cardinality of the action space. Although this does not reduce the theoretical computational complexity of MCSA, in practice, the computation time may reduce. MCSA requires to store the values of $A_H; D; B_H$ and O_L. However, it need not store any information regarding the state space. Hence, the storage complexity of MCSA is $O(1)$.

The per-iteration worst case computational complexity of SMCSA is also $O(jAj)$ because when the current values of the constraints are below the specified values, the associated procedures are same as that of MCSA. However, the complexity involved with the probabilistic state-aware blocking and offloading is $O(1)$ because no comparison among the actions is required. Apart from A_H, D, B_H and O_L, SMCSA needs to store the information regarding the regions $R_n(1 \quad n \quad P)$ and corresponding probabilities $q(n)$ and $p(n)$. Therefore, the storage complexity of SMCSA is $O(P)$. Clearly, storage complexities of the proposed algorithms are significantly lower than those of learning schemes [7], [8] having a storage complexity of at least $O(jSj)$.

Table 3: Complexities of different algorithms.

Algorithm	Storage complexity	Computational complexity
MCSA	$O(1)$	$O(\lvert\mathcal{A}\rvert)$
SMCSA	$O(P)$	$O(\lvert\mathcal{A}\rvert)$

Remark 9. *A well-studied approach for MDP problems is the investigation of structural properties, see e.g., [6], [48], which often leads to a threshold-based optimal policy. In [6], which does not consider offloading probability of low priority users, we prove the optimality of threshold policies. Although the computational complexity of the resulting algorithm in [6] is lower than that of policy iteration, it is still exponential in one of the parameters of the state space. The problem addressed in this paper does not*

result in a threshold-based optimal policy. However, the proposed algorithms provide significantly lower computational complexities compared to what would have been achieved corresponding to a threshold-based optimal policy.

6 Simulation Results

In this section, we evaluate the performances of the proposed algorithms in ns-3. We utilize the open source ns-3 simulation package for 5G NR developed in [11], [12]. The simulation package consists of customizable PHY and MAC layers of mmWave based 5G NR stack. Higher layers of the 5G stack are implemented following the design principles of corresponding LTE stack in ns-3 [13]. One can tune various 5G system parameters such as bandwidth, frame structure within the simulation package. We observe performances of the proposed algorithms and the optimal policy in terms of the blocking probability of high priority users, the offloading probability of low priority users and the total system throughput. We also compare the performances of the proposed algorithms with the association scheme adopted in existing network. In this scenario, the association scheme results in on-the-spot offloading [49], where low priority users are always associated with WiFi and high priority users are associated with 5G NR. However, when capacity is reached in 5G NR, high priority users are blocked. We also compare the performances of our proposed algorithms with on-the-spot offloading in the face of user mobility.

6.1 Simulation Setup and Methodology

Our simulation setup consists of a centralized controller which comprises a 5G NR controller and a WiFi controller as logical entities. RAT selection functionalities are handled in the centralized controller. The network model is composed of a 3GPP 5G NR gNB and an operator-deployed IEEE 802.11g WiFi AP inside the coverage area of the 5G NR gNB. Users are assumed to be stationary. We set the radius of the coverage area of WiFi AP to be around 30 m. WiFi AP is located at nearly 50 m from 5G NR gNB. 5G NR and WiFi parameters are described in Tables 4 and 5. 5G NR and WiFi parameters are selected based on 3GPP [39] models (rural and low density urban macro scenario) and saturation throughput [40] IEEE 802.11g WiFi [50] model, respectively. The carrier frequency for 5G NR is chosen to be 700 MHz, and the bandwidth is 20 MHz [39]. In simulations, we assume that a low priority user can obtain a maximum data rate of 10 Mbps due to access network bottleneck. We set $B_{\max} = O_{\max} = 0.05$, $_B = _O = 0.01$. In case of SMCSA, we divide the entire state space into two regions, viz., R_1 and R_2. We keep $q(1) = p(1) = 0$ and $q(2) = p(2) = 1$. We choose $d = p_c = 2$.

Remark 10. *We consider that high and low priority users use VoIP and video services, respectively. Note that the bit rate and the packet payload for high priority VoIP users (see Table 4) are chosen in accordance with [51].*

6.2 High Priority User Arrival Rate Variation

Fig. 3a describes the high priority user blocking probability performances of the proposed algorithms, optimal policy (without the consideration of multiple user offloading) and on-the-spot offloading (existing RAT selection algorithm) as a function of $_H$. The blocking probability of the optimal policy increases with $_H$. Since MCSA blocks high priority users based on $B_{\max}$ (irrespective of $_H$), the blocking probability is nearly the same for all $_H$s. SMCSA is designed in such a way that it blocks high priority users only when the system reaches region R_2. We consider two cases, viz., $R_1 : (i_G + 2i_B) \quad C_N \quad 2$, and $R_1 : (i_G + 2i_B) \quad C_N \quad 1$, respectively. The high priority user blocking probability of SMCSA gradually increases with $_H$, similar to the optimal policy. This happens because when the value of $_H$ is low, we block the incoming high priority users with low probability. As $_H$ increases and the system gradually fills up with high priority users, the probability of blocking increases as $q(n)$ increases with n. Since the size of the region R_1 is smaller in the first case, the blocking probability is lower in the second case. In case of on-the-spot offloading, the high priority user blocking probability gradually rises with $_H$. Since blocking happens only when the system reaches capacity, the blocking probability is lower than those of other algorithms.

In Fig. 3b, we plot the low priority user offloading fractions for the considered algorithms. In existing network, offloading is never performed for the considered scenario. The low priority user offloading probabilities of MCSA and SMCSA are similar for all values of $_H$ since $_L$ is fixed. Changes in $_H$ do not have much impact on the offloading probability of low priority users. However, the low priority user offloading probability of the optimal policy rises with $_H$ because with increasing $_H$, actions involving offloading $(A_4; A_5; A_6; A_7)$ are selected more frequently.

The total system throughput provided by MCSA is very close to the total system throughput of the optimal policy,

Table 4: 5G NR Network Model.

Parameter	Value
High priority user capacity	4 users
Bit rate of a high priority user	20 kbps
High priority user packet payload	50 bytes
Low priority user packet payload	600 bytes
Tx power for gNB and MS	49 dBm and 23 dBm
Noise figure for gNB and MS	5 dB and 9 dB
Antenna height for gNB and MS	35 m and 1.5 m
Path loss (R in kms)	$128.1 + 37.6 \log(R)$
Multi-path fading	Extended Pedestrian A model [52]

Table 5: WiFi Network Model.

Parameter	Value
Channel bit rate	54 Mbps
UDP header	224 bits
Packet payload	1500 bytes
Slot duration	20 s
Short inter-frame space (SIFS)	10 s
Distributed Coordination Function IFS (DIFS)	50 s
Minimum acceptable per-user throughput	4.5 Mbps
Tx power for AP	23dBm
Noise figure for AP	4 dB
Antenna height for AP	2.5 m
Antenna parameter	Isotropic antenna
Path loss (R in kms)	$140.3 + 36.7 \log(R)$
Fading	Rayleigh fading

(a) High priority user blocking percentage vs. $_H$

(b) Low priority user of oading percentage vs. $_H$

(c) Total system throughput vs. $_H$.

(d) High priority user blocking percentage vs. $_L$

(e) Low priority user of oading percentage vs. $_L$

(f) Total system throughput vs. $_L$.

Figure 3: Plot of different system parameters for different algorithms under varying $_H$ ($_L = 1$; $_H = 1$ and $_L = 1$) and varying $_L$ ($_H = 0:2$; $_H = 1$ and $_L = 1$).

as observed in Fig. 3c. The total throughput of SMCSA in both the cases are slightly lower than that of MCSA because MCSA blocks more fraction of high priority users than those by SMCSA. Since there is a trade-off between the total system throughput and the high priority user blocking probability, the total system throughput is higher in the case of MCSA. All these algorithms perform better than on-the-spot of oading. Since in on-the-spot of oading, low priority users are always associated with WiFi, no load balancing mechanism is present. Moreover, the total throughput in WiFi degrades due to contention among low priority users. Hence, the throughput performance of on-the-spot of oading is the worst among all.

6.3 Low Priority User Arrival Rate Variation

Fig. 3d illustrates the high priority user blocking probability performances of different algorithms as a function of $_L$. Similar to Fig. 3a, the blocking probability of MCSA is close to B_{max} for all $_L$. SMCSA blocks high priority users and of oads low priority users only when the system reaches region R_2. We consider two cases, viz., $R_1 : (i_G + 2i_B)$ C_N $2; (k_G + k_B)$ $4; (j_G + j_B)$ 2 and $R_1 : (i_G + 2i_B)$ C_N $2; (k_G + k_B)$ $4; (j_G + j_B) < 2$, respectively. The performance of SMCSA for the rst case is close to that of the optimal policy. In the second case, the blocking probabilities are slightly higher than those of the rst case because R_1 is smaller in the second case. Since on-the-spot of oading always associates high priority users with 5G NR, changes in $_L$ do not affect the blocking probability of high priority users.

In Fig. 3e, the of oading probability of the optimal policy grows with $_L$. The of oading probability of MCSA is close to O_{max} for every $_L$. In SMCSA, the of oading probability grows with $_L$ (similar to the optimal policy) because $p(n)$ is an increasing function of n. The of oading probability in the second case is slightly larger than in the rst case since region R_1 is smaller in the second case. Since of oading is not possible in the case of on-the-spot of oading, the of oading probability of low priority users is always zero.

In Fig. 3f, we observe that the performances of both MCSA and SMCSA are close to optimal, outperforming on-the-spot of oading algorithm. Though the proposed algorithms take into account only the instantaneous rewards while optimizing, these algorithms facilitate load balancing between 5G NR and WiFi. The total system throughput of on-the-spot of oading does not increase much with $_L$ due to contention among users in WiFi. Fig. 3f demonstrates that indeed our proposed algorithms provide near-optimal performances. The gain in total throughput obtained by the proposed algorithms with respect to on-the-spot of oading is more in Fig. 3f compared to Fig. 3c because the low priority users have a higher contribution to the total system throughput than high priority users. The throughput of a high priority user (20 kbps, see Table 4) is signi cantly

smaller than the throughput of a low priority user in 5G NR (remaining resources are allocated uniformly) and WiFi (minimum per-user throughput=4.5 Mbps, See Table 5), depending on the number of users in the system.

6.4 Consideration of Realistic 5G NR Capacity

In previous simulations, to compare the performances of the proposed algorithms with the optimal policy, we set the high priority user capacity in 5G NR to a small value. This is needed for the computation of the optimal policy for which the computational complexity becomes intractable, even for moderate values of system parameters.

In this section, we set the high priority user capacity in 5G NR to 40 and demonstrate that the proposed algorithms perform better than on-the-spot of oading for both under-load (low arrival rates of high and low priority users) and overload scenarios (high arrival rates of high and low priority users). For SMCSA, we consider $R_1 : (i_G + 2i_B) < 30; (k_G + k_B) \quad 4; (j_G + j_B) < 30$. As demonstrated in Fig 4a and 4b, in under-load condition, both MCSA and SMCSA perform better than on-the-spot of oading. However, the throughput obtained by MCSA is slightly higher than that of SMCSA since SMCSA blocks and of oads based on the system state. Similar observation is made (Fig. 4c and 4d) in overload condition when the arrival rates are high. We have not shown the blocking probability and the of oading probability performances since they follow similar trend as those of Figs 3a, 3d, 3b and 3e.

6.5 Consideration of Multiple Channel States in WiFi

In the system model, we have assumed that channel states of users in WiFi are always good. In this section, we evaluate the performance of our proposed algorithms considering multiple channel states of users in WiFi. We consider two types of low priority users. As assumed previously, users present within the coverage area of the WiFi AP are taken to be users with good channel states in WiFi. Users present outside the coverage area of the WiFi AP are assumed to be users with bad channel states in WiFi. Such users are always associated with 5G NR, irrespective of their channel states in 5G NR.

In Fig. 5a, we plot the total system throughputs for the considered algorithms as a function of $_H$. As observed from the gure, both the proposed algorithms perform better than on-the-spot of oading. The throughput performance of MCSA is slightly better than that of SMCSA. However, the performances of SMCSA and on-the-spot of-oading are very close to each other. This is due to the fact that for users with bad channel in WiFi, both on-the-spot of oading and the proposed algorithms work in a similar fashion since all these algorithms associate them with the 5G NR gNB. Therefore, performance bene ts corresponding to the proposed algorithms are achieved only due to users which do not have bad channels with respect to the WiFi AP. Since separate resources are reserved in 5G NR for low priority users outside the coverage area of the WiFi AP, the blocking probabilities and the of oading probabilities are identical to those of Figs. 3a and 3b.

Similarly, in Fig. 5b, we illustrate the comparative performances of MCSA, SMCSA and on-the-spot of oading in terms of the total system throughput as a function of $_L$. Clearly, both MCSA and SMCSA outperform on-the-spot of-oading. As $_L$ increases, the performance gap between the proposed algorithms and on-the-spot of oading, increases. The blocking probabilities and the of oading probabilities are identical to those of Figs. 3d and 3e.

6.6 Consideration of CQI in 5G NR

In this section, we propose and evaluate the performances of variants of MCSA and SMCSA so as to take into account 16 CQI values in 5G NR standardized by 3GPP and call them MCSA-c and SMCSA-c, respectively. The modi cation to MCSA and SMCSA is as follows. While choosing an action involving of oading of a low priority user ($A_4; A_5; A_6$ and A_7), we always choose the user with the lowest CQI value in 5G NR for of oading to WiFi and the user with the lowest SNR in WiFi for of oading to 5G NR. For example, when A_4 is chosen, we choose the low priority user with the worst CQI among the users with bad channels in 5G NR, for of oading to WiFi. Since a user with bad CQI provides low throughput, we choose the user with the worst CQI for of oading. Fig. 6a and 6b illustrate that both MCSA-c and SMCSA-c outperform on-the-spot of oading in terms of total system throughput. The blocking probabilities and the of oading probabilities are identical to those of Figs. 3a, 3d, 3b and 3e.

6.7 Consideration of Mobility

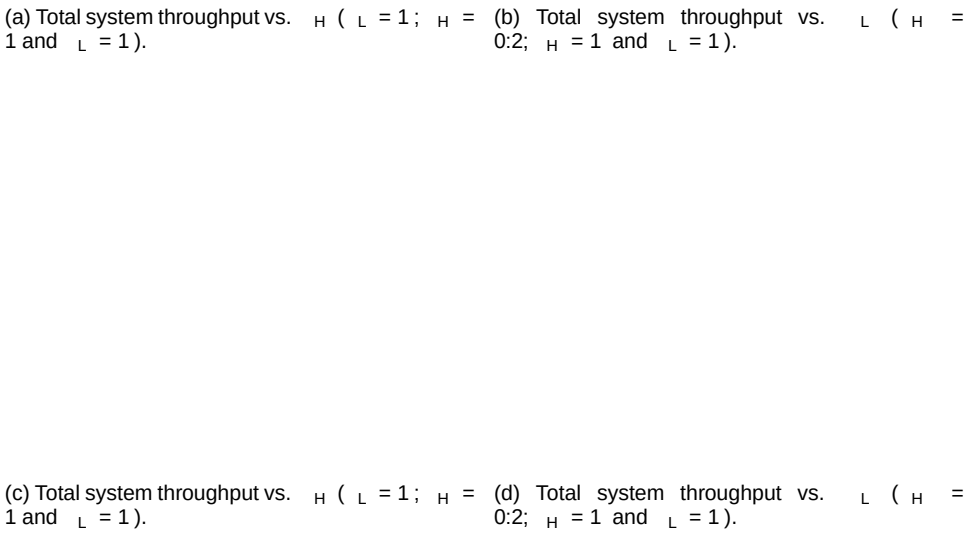
In this section, we evaluate the performances of the algorithms in the presence of user mobility. We consider random waypoint model [53] for user mobility. We set the user speed in the range $[0; 40]$ km/h.

Mobile users may be of oaded frequently from one RAT to another. This may increase the of oading probability of the overall system. Since the proposed algorithms are designed in such a way that the of oading probability satis es the constraint, a mobile user may signi cantly increase the of oading probability of the system. As a result, it may happen that stationary users get very less number of of oading opportunities. Furthermore, a user with mobility is expected to drain a lot of battery due to excessive of oading from one RAT to another. To take into account these factors, we modify the algorithms in the following way. Apart from the constraint on the overall of oading probability of low priority users, we consider of oading pro le of individual low priority users while of oading. To be precise, whenever an action involving of oading of low priority users ($A_4; A_5; A_6$ and A_7) is chosen, we choose a user which has not been of oaded till now. If no such user is present, then we choose the user which has been of oaded the earliest before.

In Fig. 7a, we observe that both MCSA and SMCSA outperform on-the-spot of oading in terms of the total system throughput. Similar observation holds in Fig. 7b for varying $_L$. We have not shown the blocking probability and the of oading probability performances since they are exactly same as those of Figs 3a, 3d, 3b and 3e.

6.7.1 Consideration of Channel States

In the presence of mobility, the channel states of users may vary over time. Hence, it may not be appropriate to



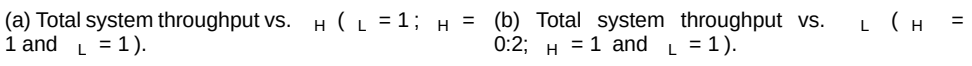
(a) Total system throughput vs. $_H$ ($_L$ = 1 ; $_H$ = 1 and $_L$ = 1).

(b) Total system throughput vs. $_L$ ($_H$ = 0:2; $_H$ = 1 and $_L$ = 1).

(c) Total system throughput vs. $_H$ ($_L$ = 1 ; $_H$ = 1 and $_L$ = 1).

(d) Total system throughput vs. $_L$ ($_H$ = 0:2; $_H$ = 1 and $_L$ = 1).

Figure 4: Plot of total system throughput for different algorithms under realistic 5G NR capacity and varying $_H$ and $_L$.

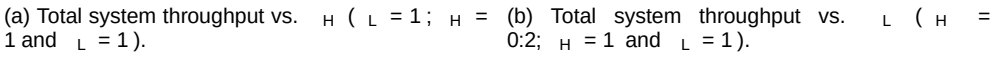


(a) Total system throughput vs. $_H$ ($_L$ = 1 ; $_H$ = 1 and $_L$ = 1).

(b) Total system throughput vs. $_L$ ($_H$ = 0:2; $_H$ = 1 and $_L$ = 1).

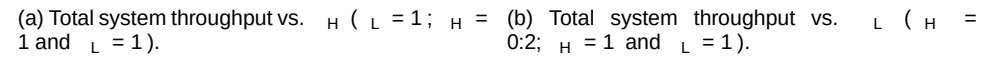
Figure 5: Total system throughput for different algorithms under varying $_H$ and $_L$.

assume that users are distributed in cell edge/cell center region depending on their average radio conditions. To take into account this factor, we propose modi cations to MCSA and SMCSA and call them MCSA-m and SMCSA-m, respectively. The modi cation is as follows. Whenever the channel state of a user changes, we update the system state. For example, when the channel state of a high priority user changes from good to bad, then we increase and decrease the number of high priority users with bad and good channels in the state space, respectively, by one. This can be viewed as if a user with good channel has departed, and a user with bad channel has arrived. Figs. 7a and 7b illustrate that MCSA-m and SMCSA-m outperform on-the-spot of oading. Moreover, MCSA-m and SMCSA-m perform marginally better than MCSA and SMCSA, respectively as we take into account the changes in channel states (due to mobility) in the state space.



(a) Total system throughput vs. $_H$ ($_L$ = 1 ; $_H$ = 1 and $_L$ = 1). (b) Total system throughput vs. $_L$ ($_H$ = 0:2; $_H$ = 1 and $_L$ = 1).

Figure 6: Total system throughput for different algorithms under consideration of CQI in 5G NR and varying $_H$ and $_L$.



(a) Total system throughput vs. $_H$ ($_L$ = 1 ; $_H$ = 1 and $_L$ = 1). (b) Total system throughput vs. $_L$ ($_H$ = 0:2; $_H$ = 1 and $_L$ = 1).

Figure 7: Plot of total system throughput for different algorithms under user mobility and varying $_H$ and $_L$

6.8 Consideration of Large Network with User Mobility

In this section, we evaluate the performances of variants of our algorithms (MCSA-l and SMCSA-l, respectively) in a large network consisting of 5 5G NR gNBs and 5 WiFi APs. The inter-gNB distance is 200 m [39], and WiFi APs are present in hotsopt regions. The users are mobile, following the model described in Section 6.7.

To take into account multiple gNBs and APs, we adopt the following strategy. As stated in Remark 8, we map every user to the gNB (AP) providing highest SNR. Also, similar to MCSA-m and SMCSA-m (Section 6.7), we take into account of oading pro le and channel states of individual users while making RAT selection and of oading decisions to handle user mobility. Figs. 8a and 8b illustrate that MCSA-l and SMCSA-l perform better than on-the-spot of oading in terms of total system throughput. Moreover, MCSA-l performs marginally better than SMCSA-l.

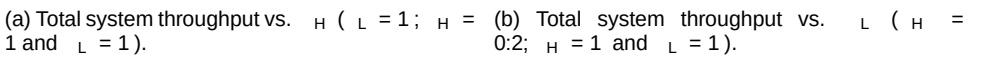
6.9 Consideration of Multiple User Of oading

As multiple user of oading may lead to signi cant instantaneous control signaling in the core network and increase in the computation complexity, the system model considered by us take into account single user of oading only. However, our model can be extended to incorporate of oading of multiple users in the action space. Of oading of multiple users may provide higher instantaneous reward than single user of oading depending on the system state. Therefore, the total system throughput corresponding to the policy involving multiple user of oading may be more than that involving single user of oading. To this end, we compare the optimal policy considering multiple user of oading as actions in the action space, with the optimal policy without considering multiple user of oading as actions in the action space (single user of oading only). As demonstrated in Fig. 9, we observe that indeed consideration of multiple user of oading results in marginal improvement in the total system throughput.

7 Discussion and Conclusion

Optimal RAT selection problem in a 5G NR-WiFi HetNet aiming to maximize the total system throughput subject to constraints on high priority user blocking probability and low priority user of oading probability is formulated as a CMDP. The key insights from our analysis are as follows.

Elimination of sub-optimal actions in different states reduces the size of the effective action space. Still, DP techniques for the computation of optimal policy suffer from the curses of dimensionality and modeling.



(a) Total system throughput vs. $_H$ ($_L$ = 1 ; $_H$ = 1 and $_L$ = 1).

(b) Total system throughput vs. $_L$ ($_H$ = 0:2; $_H$ = 1 and $_L$ = 1).

Figure 8: Total system throughput for different algorithms in a large network with user mobility and varying $_H$ and $_L$.

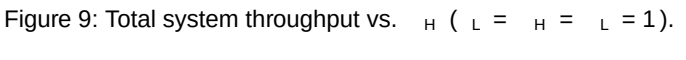


Figure 9: Total system throughput vs. $_H$ ($_L$ = $_H$ = $_L$ = 1).

Similar to learning based approaches, both MCSA and SMCSA are free from the curse of modeling. Although optimal performance may not be guaranteed, our schemes do not suffer from high storage complexity and slow convergence issues present in learning based schemes. Contrary to MCSA, SMCSA performs blocking and of- oading based on the system state and hence, is closer in spirit to the optimal policy. The per-iteration computational complexity of both algorithms is O(jAj). Storage complexities of MCSA and SMCSA are O(1) and O(P) since SMCSA needs to store additional information regarding the regions R_n. Thus, the proposed algorithms overcome curses of dimesnionality and modeling associated with the DP based methods to compute the optimal policy.

Simulations conducted in ns-3 based 5G NR HetNet indicate that both MCSA and SMCSA outperform traditional algorithms under various network scenarios.

To summarize, in this paper, we consider the optimal RAT selection problem in a 5G NR-WiFi HetNet consisting of users of multiple priorities and channel states. Maximizing the total system throughput subject to constraints on the high priority user blocking probability and the low priority user of oading probability is formulated as a CMDP problem. We prove the sub-optimality of different actions in different states. We then propose two low-complexity algorithms for RAT selection in a 5G NR-WiFi network. These algorithms do not require the knowledge of the unknown system dynamics. Contrary to the rst algorithm where the blocking probability and the of oading probability do not depend on the statistics of the arrival processes, in the second algorithm, blocking and of oading are performed based on the system state. Simulation conducted in a ns-3 based 5G NR-WiFi HetNet establish that the proposed algorithms outperform traditional algorithms, even in the face of user mobility.

8 PROOFS

8.1 Proof of Lemma 1

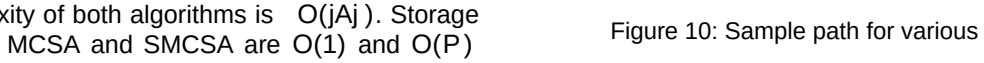


Figure 10: Sample path for various policies.

We prove the lemma using sample path arguments. We consider case of event E_5. Proofs for the other events follow in a similar manner. We assume that the system starts at time $t = 0$. Suppose that the system is in state $s_1 = (i_G; i_B; j_G; j_B; k_G; k_B)$, when event E_5 occurs at time t_1. Consider a policy which chooses A_7 in state $(i_G; i_B; j_G; j_B; k_G; k_B)$ and denote it by $_1$. Consider another policy (may be a non-stationary policy) $_2$ which selects A_5 in state $(i_G; i_B; j_G; j_B; k_G; k_B)$. As illustrated in Fig. 10, let us assume that under policies $_1$ and $_2$, the system move from state s_1 to state $s_2 = (i_G; i_B; j_G; j_B + 1; k_G; k_B - 1)$ and $s_3 = (i_G; i_B; j_G + 1; j_B; k_G - 1; k_B)$, respectively. Since we consider a Markovian system, the inter-arrival and service times are identical for the considered sample paths. We assume that based on the next event E_l and following

the policy $_1$, the system moves from state s_2 to state s_4. Suppose the policy $_2$ is such that in response to event E_l, it chooses the same action as that of $_1$. Additionally, it of oads one good user from 5G NR to WiFi and one bad user from WiFi to 5G NR. Therefore, under the policy $_2$, the system moves from state s_3 to s_4. We construct the policy $_2$ in such a way that here onwards, it chooses the same action as that of policy $_1$. Therefore, from state s_4 onwards, both the sample paths follow the same trajectory. The difference of value functions of state s_1 under policies $_1$ and $_2$ is given by

$$V_{_1}(s_1) \quad V_{_2}(s_1) = \frac{(C_N \quad i_G \quad p_c i_B)}{(j_G + j_B + 1)} R_{L;L} \left(\frac{1}{d} \quad 1\right) < 0:$$

Therefore, policy $_2$ is strictly better than $_1$. Since the Markov chains under various policies are recurrent in nature, state s_1 is visited in nitely often. Upon each visit, action A_7 induced by policy $_1$ provides lesser reward than action A_5 corresponding to $_2$. This completes the proof of the lemma.

8.2 Proof of Lemma 2

Similar to Lemma 1, we prove this lemma for event E_9. Proof for event E_{10} follows in a similar way. Suppose the system is in state $s_1 = (i_G; i_B; j_G; j_B; k_G; k_B)$ when event E_9 occurs at time t_1. Consider policies $_1$ and $_2$ which choose A_5 and A_7 in state s_1, respectively. We assume that under policies $_1$ and $_2$, the system move from state s_1 to state $s_2 = (i_G; i_B; j_G; j_B \quad 1; k_G; k_B + 1)$ and $s_3 = (i_G; i_B; j_G \quad 1; j_B; k_G + 1; k_B)$, respectively. We assume that based on the next event E_l and following the policy $_1$, the system moves from state s_2 to state s_4. Suppose the policy $_2$ is such that for event E_l, it chooses the same action as that of $_1$, of oads one bad user from 5G NR to WiFi and one good user from WiFi to 5G NR. Therefore, under the policy $_2$, the system moves from state s_3 to s_4. We construct the policy $_2$ in such a way that here onwards, it chooses the same action as that of policy $_1$. Therefore, from state s_4 onwards, both the sample paths follow the same trajectory. The difference of value functions of state s_1 under policies $_1$ and $_2$ is given by

$$V_{_1}(s_1) \quad V_{_2}(s_1) = \frac{(C_N \quad i_G \quad p_c i_B)}{(j_G + j_B + 1)} R_{L;L} \left(1 \quad \frac{1}{d}\right) > 0:$$

Therefore, policy $_1$ is strictly better than $_2$. Due to the recurrent nature of the Markov chain, similar to Lemma 1, A_5 is better than A_7.

ACKNOWLEDGMENT

This work has been supported by the Department of Telecommunications, Ministry of Communications, India as part of the Indigenous 5G Test Bed project.

REFERENCES

[1] A. Roy, P. Chaporkar, A. Karandikar, and P. Jha, "Optimal radio access technology selection in an SDN based LTE-WiFi network," in IEEE WiOpt RAWNET Workshop, 2019, pp. 1–8.

[2] 3GPP TS 23.501 v16.6.0, "System Architecture for the 5G System," Dec. 2020, [Online]. Available: https://portal.3gpp.org/desktopmodules/Speci cations/Speci cationDetails.aspx?speci cationId=3144.

[3] 3GPP TS 38.401, "Next Generation Radio Access Network (NG-RAN) Architecture Description," 2017, [Online]. Available: https://portal.3gpp.org/desktopmodules/Speci cations/Speci cationDetails.aspx?speci cationId=3219.

[4] A. N. Manjeshwar, A. Roy, P. Jha, and A. Karandikar, "Control and management of multiple RATs in wireless networks: An SDN approach," in IEEE 5GWF, 2019, pp. 596–601.

[5] A. N. Manjeshwar, P. Jha, and A. Karandikar, "A centralized SDN architecture for the 5G cellular network," in IEEE 5GWF, 2018, pp. 147–152.

[6] A. Roy, P. Chaporkar, and A. Karandikar, "Optimal radio access technology selection algorithm for LTE-WiFi network," IEEE Transactions on Vehicular Technology, vol. 67, no. 7, pp. 6446–6460, 2018.

[7] ——, "An on-line radio access technology selection algorithm in an LTE-WiFi network," in IEEE WCNC, 2017, pp. 1–6.

[8] A. Roy, V. Borkar, P. Chaporkar, and A. Karandikar, "Low complexity online radio access technology selection algorithm in LTE-WiFi HetNet," IEEE Transactions on Mobile Computing, vol. 19, no. 2, pp. 376–389, 2019.

[9] 3GPP TR 37.834 v0.3.0, "Study on WLAN/3GPP Radio Interworking," 2013, [Online]. Available: https://portal.3gpp.org/desktopmodules/Speci cations/Speci cationDetails.aspx?speci cationId=2623.

[10] "Network simulator-3," [Online]. Available: http://code.nsnam.org/ns-3-dev/.

[11] M. Mezzavilla, M. Zhang, M. Polese, R. Ford, S. Dutta, S. Rangan, and M. Zorzi, "End-to-end simulation of 5G mmwave networks," IEEE Communications Surveys & Tutorials, vol. 20, no. 3, pp. 2237–2263, 2018.

[12] M. Mezzavilla, S. Dutta, M. Zhang, M. R. Akdeniz, and S. Rangan, "5G mmwave module for the ns-3 network simulator," in ACM MSWiM, 2015, pp. 283–290.

[13] "The LENA ns-3 LTE module documentation," [Online]. Available: http://iptechwiki.cttc.es/LTE-EPC _Network _Simulator _(LENA).

[14] M. Singh and P. Chaporkar, "An ef cient and decentralised user association scheme for multiple technology networks," in IEEE WiOpt, 2013, pp. 460–467.

[15] A. Whittier, P. Kulkarni, F. Cao, and S. Armour, "Mobile data of oading addressing the service quality vs. resource utilisation dilemma," in IEEE PIMRC, 2016, pp. 1–6.

[16] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data of oading: How much can WiFi deliver?" IEEE/ACM Transactions on Networking, vol. 21, no. 2, pp. 536–550, 2013.

[17] D. Suh, H. Ko, and S. Pack, "Ef ciency analysis of WiFi of oading techniques," IEEE Transactions on Vehicular Technology, vol. 65, no. 5, pp. 3813–3817, 2016.

[18] N. Cheng, N. Lu, N. Zhang, X. Zhang, X. S. Shen, and J. W. Mark, "Opportunistic WiFi of oading in vehicular environment: A game-theory approach," IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 7, pp. 1944–1955, 2016.

[19] Q. Ye, B. Rong, Y. Chen, M. Al Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," IEEE Transactions on Wireless Communications, vol. 12, no. 6, pp. 2706–2716, 2013.

[20] P. Naghavi, S. H. Rastegar, V. Shah-Mansouri, and H. Kebriaei, "Learning RAT selection game in 5G heterogeneous networks," IEEE Wireless Communications Letters, vol. 5, no. 1, pp. 52–55, 2015.

[21] D. D. Nguyen, H. X. Nguyen, and L. B. White, "Reinforcement learning with network-assisted feedback for heterogeneous RAT selection," IEEE Transactions on Wireless Communications, vol. 16, no. 9, pp. 6062–6076, 2017.

[22] W. Song, Y. Cheng, and W. Zhuang, "Improving voice and data services in cellular/WLAN integrated networks by admission control," IEEE Transactions on Wireless Communications, vol. 6, no. 11, pp. 4025–4037, 2007.

[23] C. Liu, M. Li, S. V. Hanly, and P. Whiting, "Joint downlink user association and interference management in two-tier HetNets with dynamic resource partitioning," IEEE Transactions on Vehicular Technology, vol. 66, no. 2, pp. 1365–1378, 2017.

[24] S. Barmpounakis, A. Kaloxylos, P. Spapis, and N. Alonistioti, "Context-aware, user-driven, network-controlled RAT selection for 5G networks," Computer Networks, vol. 113, pp. 124–147, 2017.

[25] W. Huang, D. Meng, J. N. Hwang, J. Park, Y. Xu, and W. Zhang, "QoE based SDN heterogeneous LTE and WLAN multi-radio networks for multi-user access," in IEEE WCNC, 2018, pp. 1–6.