MATCH: Multiple Access for multiple Traffic Classes in 5G HetNets

Virgilios Passas†*, Nikos Makris†*, Vasileios Miliotis†*, Thanasis Korakis†* and Leandros Tassiulas†
†Department of Electrical and Computer Engineering, University of Thessaly, Greece
*Centre for Research and Technology Hellas,CERTH, Greece
‡Department of Electrical Engineering, Yale University, New Haven, USA
Email: {vipassas, nimakris, vmiliotis, korakis}@auth.gr, leandros.tassiulas@yale.edu

Abstract—Ultra-Dense Heterogeneous Network deployments are expected to boost the offered network capacity and enhance the user-perceived Quality of Experience, through the simultaneous offering of multiple technologies using distinct or shared wireless spectrum. In such environments with a plethora of available Radio Access Technologies (RATs), the network UEs shall decide either independently or assisted through operator based services on which network they shall use to better serve their needs. In this work, we model the network selection problem in a Multi-RAT system, based on the Paris Metro Pricing (PMP) scheme, enhanced with dynamic pricing formed by the congestion of each available technology. We assume that the network UEs are equipped with multi-homing features and thus are able to use concurrently more than one technologies, based on the requirements of the applications requesting network connectivity (e.g. UHD video streaming). We port and experimentally evaluate the proposed system model over a testbed setup, using distributed components running at the UEs and at a Core Network controller. We provide evaluation results on the average cost per UE for the selected RATs, the average data rate distribution of each UE per RAT and how these performance metrics are affected under different client ordering policies at the network controller.

I. INTRODUCTION

Ultra Dense Heterogeneous Networks (HetNets) are expected to bring numerous advantages for the mobile network operators (MNOs) of the next generation cellular networks, by offering enhanced network capacity and diverse technologies for serving the end-users. Through enhanced spectral techniques and spectrum coordination among the available cells, 5G networks will utilize bands ranging from the sub6-GHz to cm- and mm-Wave bands. Yet, this type of equipment upgrade for the Radio Access Network (RAN) incurs extra deployment and operation costs for the MNOs. When considering that the Average Revenue Per User (ARPU) is either flat or even slowly decreasing [1], it becomes clear that cost efficient techniques are required in order to keep in pace with the demand-driven evolution in the cellular technology.

Nevertheless, through the utilization of off-the-shelf components operating in multiple bands and with different technologies, HetNets can be formed with low CAPEX and OPEX costs. Forming HetNets for adding to the overall network capacity has existed for legacy technologies as well, (e.g. UMTS, LTE) through the addition of low energy consuming devices with lower coverage in the network (e.g. femto-/pico-cells). Recent efforts have also focused on the inter-networking of cellular technologies with WiFi as well, as a means to add-up to the offered network capacity with low-cost solutions (e.g. through the LTE-WiFi Aggregation Adaptation Protocol [2]).

The incorporation of the new spectrum that 5G is expected to operate, creates a complex ecosystem for allocating the operator served clients to each of the offered technologies, while meeting diverse demands for network capacity.

This technology allocation problem becomes even more complicated when considering that the perceived Quality of Experience (QoE) highly depends on the application that is running on the network UE. Different applications have versatile demands for the network capacity; UHD video requires more than 25Mbps to be dedicated for the application, whereas an email application is only served as background traffic. Based on the existing technology offering for running multi-homed end-clients (e.g. MPTCP, using SDN and software switches [3]), a network client might benefit from having simultaneously several connections of different technologies active to the Internet. In such an environment, efficient mapping of each application to the available technologies might significantly enhance the perceived QoE, whereas alleviate the burden placed on the operator serving multiple data-intensive traffic streams. Nevertheless, multiple traffic classes for the applications add to the complexity of the network resources allocation to the served clients.

The main questions that arise in such heterogeneous environments with multiple RATs being offered to the network UEs, and multiple traffic classes per UE are:

1) Which technology should a UE select for the different types of traffic.
2) When should a UE switch the serving network for a specific traffic class.
3) How should an operator charge each RAT.
4) How these decisions affect the overall system stability.

In order to answer these questions, we extend our previous work in [4], where we introduced a network selection scheme based on the Paris Metro Pricing (PMP), a service differentiation scheme that was first used in Paris metro to give to its passengers the ability to opt for less congested wagons. In this paper we propose a system model that takes into consideration multiple traffic classes concurrently utilized by each UE, and multiple networks being offered by the MNO, forming a Multi-

Classes in 5G HetNets
RAT system. We evaluate the performance and stability of the system through extensive experimentation over multiple technologies in an open wireless testbed.

The rest of the paper is organized as follows: Section II presents relevant works on the RAT selection policies and methodologies. In Section III we introduce our system model. Section IV includes the description of our system architecture, and in Section V we showcase our experimental findings. Finally, in Section VI we conclude our work.

II. RELATED WORK

Different access schemes have been proposed in literature for exploiting the coexistence of heterogeneous wireless technologies for improving the Quality of Service (QoS) and Experience (QoE) of the mobile users. In [5], a survey on the different models for network selection in HetNets is provided, with the solutions being classified based on the proposed utility functions and system models. Similarly, in [6], authors classify the respective algorithms based on the location where the decision making network selection components are running, as either partially or fully distributed. In [7], a fully distributed access point selection algorithm is presented based on no-regret learning. Through the application of this scheme, the system is able to reach a correlated equilibrium state. Authors in [8], formulate the RAT selection problem using a dynamic evolutionary game and introduce a centralized algorithm based on reinforcement learning. In [9], the RAT selection problem is modeled as a non-cooperative game, and is evaluated for its convergence, efficiency, and practicality. Through their approach, each user tries to selfishly maximize its own throughput, while the impact of a user’s decisions on other users performance and the convergence of the system to Nash equilibria is investigated. The authors conclude that an improvement path can be repeated infinitely with a mixture of classes. As network convergence is the target of these algorithms, authors in [10] discuss the convergence properties of network selection games. The network selection process is studied as a non-cooperative game, and is evaluated for the cases where each client uses its own preference to select a network, and for a combination of client and network preferences to arrive at pairings.

Contrary to the majority of works that deal with the maximization of the user throughput, authors in [11] propose the user demand-centric optimization, where users seek to maximize quality of experience (QoE). Their research validates the existence of user demand diversity gain and the effectiveness of their learning algorithm in improving the system efficiency and QoE fairness. Authors in [12] consider a heterogeneous cellular network where each user chooses among multiple access technologies. The competition of the users is modeled as an incomplete information game where players are not aware of other players actions. An incentive mechanism that aims to motivate WiFi Access Points (APs) to participate in heterogeneous networks, by providing an access class to the existing cellular infrastructure, is proposed in [13]. The pricing strategy for the inclusion of third party WiFi APs is formulated as a Stackelberg game between the mobile network provider and the third party WiFi APs.

The authors in [14] consider a general model of congestion externality for the PMP and investigate the conditions of congestion functions that guarantee the viability of the PMP scheme. Similarly, in [4] we propose and evaluate a dynamic pricing algorithm for HetNets, based on the PMP scheme. In this work, we build on our prior experience and further extend it to MATCH (Multiple Access for multiple Traffic Classes in 5G HetNets), aiming to include multiple traffic classes per each UE, corresponding to the diverse network communication needs that applications serving the end-user are requiring. We further extend the system model with multi-homed UEs and with the inclusion of higher throughput 5G technologies (e.g. WiGig) and evaluate the proposed pricing scheme with a real testbed evaluation.

III. SYSTEM MODEL

We consider a heterogeneous environment with $M$ classes of available radio access technologies that belong to the same cellular service provider. Each class $m$ refers to a different radio access technology (e.g. 4G, 3G, WiFi, mmWave) with capacity $C_m$, resulting to a total system capacity equal to $C = \sum_{m=1}^{M} C_m$. We assume that there are $L$ types of user traffic classes and each UE $i$ may have traffic demands for these types of traffic. The traffic of each UE $i$ is characterized by the vector $\theta_i = \{\theta_{i1}, ..., \theta_{iL}\}$, where $\theta_{il} = r_{il}/w_i$ represents the sensitivity of UE$_i$’s traffic of type $l$, with $r_{il}$ being the data rate demand of UE$_i$ for its traffic class $l$, and $w_i$ the normalized spectrum efficiency of UE$_i$, with $w_i \in (0,1]$ for $i = (1, ..., N)$, which is used to abstract the physical layer for the channels of the available RATs, including the frequency selectivity due to transmissions in different frequency bands. As we focus on access layer decisions, we provide $w_i$ as a plug-in parameter to our dynamic pricing scheme, available for physical layer analysis. $\theta_{il}$ represents the ability of UE$_i$’s traffic of type $l$ to adapt easily to changes in the network conditions, while still meeting specific QoS requirements and maintaining the QoE for the user. The mass of traffic classes allocated to RAT class $m$ is denoted by $Q_m$, and without loss of generality we assume that the total mass of traffic classes of the UEs is equal to 1. The mass of traffic classes that are not allocated to a RAT class is equal to $Q_0 = 1 - \sum_{m=1}^{M} Q_m$.

We focus on UEs that are under the coverage of all provided radio access technologies as depicted in Fig. 2. Each traffic class $l$ of UE$_i$ is allocated to a RAT class $m$ such that the corresponding element $b_{im,l}^{\theta_i}$ of an $M \times L$ matrix $B_i$ is equal to $\theta_{il}$, if its traffic class $l$ is allocated to the RAT class $m$, and 0 otherwise. Thus, the traffic allocation matrix $B_i$ of UE$_i$, with columns representing the UE’s traffic classes and lines the available RAT classes, is expressed as:

$$B_i = \begin{pmatrix} \theta_{i1} & 0 & 0 & \cdots & 0 \\ 0 & \theta_{i2} & \theta_{i3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \theta_{iL} \end{pmatrix}$$
We define as $b_i$, a vector of size $1 \times M$, where each element $b_i^m$ is the sum of each row of $B_i$. Thus, $b_i^m = \sum_{l=1}^{L} b_{i,l}^m$. We also define $\alpha_i$, a RAT access index vector of size $1 \times M$, where $\alpha_i^m = 1$ if $b_i^m > 0$ and $\alpha_i^m = 0$ if $b_i^m = 0$.

A UE, with congestion sensitivity $\theta_i$, using the access classes indicated by $\alpha_i$ has a utility equal to:

$$U_i(B_i) = V(1 \cdot \alpha_i) - \alpha_i \cdot p - \theta_i \cdot f$$  \hspace{1cm} (2)

where $V$ is a flat-rate valuation of accessing a class of the multi-RAT service and $1 \cdot \alpha_i$ is the number of concurrently accessed RAT classes by UE $i$. The access charge vector $p = \{p_1, ..., p_M\}$ represents the price of each RAT at the time a UE is deciding to which RAT classes it will allocate its traffic classes. The vector of functions $f = \{f_1(Q_1, \bar{C}_1), ..., f_M(Q_M, \bar{C}_M)\}$, represents the perceived congestion of the $M$ available RAT classes by all UEs in the HetNet. Thus, $f_m(Q_m, \bar{C}_m)$ is a function for the perceived congestion at class $m$, and the available capacity of access class $m$ is denoted by $\bar{C}_m$. We assume that $V \geq p_1 > p_2 > ... > p_M$ and therefore a UE with traffic classes of sensitivity $\theta_i$ will choose, in a decomposed approach for each traffic class $l$, the service class that maximizes its discrete traffic class utility:

$$U_i^l(m) = V - p_m - \theta_i f_m(Q_m, \bar{C}_m)$$ \hspace{1cm} (3)

such that:

$$m(\theta_i) = \arg\max_{1 \leq m \leq M} U_i^l(m)$$ \hspace{1cm} (4)

or no service if $U_i^l(m) = 0$, $\forall m \in \{1, ..., M\}$ and $\forall l \in \{1, ..., L\}$. This leads to a two-level (Stackelberg) game, where the provider first decides the prices per access class, as a function of the already allocated capacity per access technology, and then the UEs distribute their traffic classes over RAT classes, selecting the most appropriate for each traffic class. The provider can play by anticipating the distribution of the UEs traffic classes over the provided service classes. The distribution of traffic classes over the provided access technologies will be a Wardrop equilibrium [15], at which no UE will have an interest in changing its traffic classes allocation to available access classes. When equilibrium of distribution of traffic classes over available access classes is reached, a given traffic class $l$ of UE $i$, of sensitivity $\theta_i$, will prefer access class $m$ over $k$ if

$$V - p_m - \theta_i f_m(Q_m, \bar{C}_m) \geq V - p_k - \theta_i f_k(Q_k, \bar{C}_k)$$ \hspace{1cm} (5)

Therefore, if $p_m - p_k \leq \theta_i (f_k(Q_k, \bar{C}_k) - f_m(Q_m, \bar{C}_m))$ and for monotonically increasing congestion perception functions $f_m(Q_m, \bar{C}_m)$, class $m$ will be preferred over $k$ if

$$\theta_i \geq (p_m - p_k)/f_m(Q_m, \bar{C}_m) - f_k(Q_k, \bar{C}_k), \text{ when } k > m$$

and if

$$\theta_i \leq (p_m - p_k)/f_m(Q_m, \bar{C}_m) - f_k(Q_k, \bar{C}_k), \text{ when } k < m$$ \hspace{1cm} (6)

This creates thresholds of $\theta$ values, $\theta_1 > \theta_2 > ... > \theta_M = \theta_{M+1} = 0$, such that at equilibrium, for $1 \leq m \leq M$, $\forall \theta_i \in (\theta_{m+1}, \theta_m)$, class $m$ is chosen, while no class is preferred for $\theta > \theta_1$. The thresholds $\theta_1, ..., \theta_{M+1}$ are defined by using the fact that, at any of these specific threshold, a UE is indifferent between choosing one of the two adjacent access classes for a specific traffic class. A traffic class, at threshold $\theta_1$, is also indifferent between using the provided service or not.

We let the congestion perception functions $f_m(Q_m, \bar{C}_m)$, $\forall m \in \{1, ..., M\}$ to be:

$$f_m(Q_m, \bar{C}_m) = \frac{Q_m}{\bar{C}_m/C_m}$$ \hspace{1cm} (7)

We introduce a dynamic pricing scheme for each access class $m$. The maximum price for each class $p_m^{\max}$ is set for accessing class $m$ when its total capacity $C_m$ is allocated, and the minimum price $p_m^{\min}$ is set when the total capacity of class $m$ is available. The price as a function of available capacity $\bar{C}_m$ is expressed in (8).

$$p_m = \max \left( p_m^{\min}, p_m^{\max} \left(1 - \frac{\bar{C}_m}{C_m} \right) \right)$$ \hspace{1cm} (8)

Regarding the mobility model of the UEs, we consider that they follow routes of diverse connectivity conditions to the available radio access technologies, inspired by the model proposed in [16]. In Fig. 1, we present the states of the Markov model for the mobility of a UE. A UE in State 0 will pass through the multi-RAT area (State 1) that we focus on with probability $p_{01}$, will stay in State 1 with probability $p_{11}$ and will leave the multi-RAT area with probability $p_{12}$. A UE starting from State 0 may not pass through the multi-RAT environment with probability $p_{02}$ and stays at State 0 with probability $p_{00}$. We consider the return probabilities from State 2 to State 1 and to State 0, equal to $p_{21}$ and $p_{10}$ respectively.

In the following sections, we provide an approach to evaluate the distributed solution we provide with MATCH. Based on the system model, we port it to run in a distributed manner on all the network, i.e. the UEs and the Core Network. Using the outcomes of the utility functions running on each network UE, a Core Network controller is able to suggest the technologies and list the prices for using them.

IV. SYSTEM ARCHITECTURE AND ALGORITHM DESIGN

For the evaluation of the proposed scheme, we use four RATs, each one with different characteristics (3G, WiFi, 4G and mmWave), and four different traffic classes available at each network UE. Each Traffic Class (TC) is categorized by its data rate demands, ranging from Traffic Class 0 (TC0) designating applications that require Best Effort connectivity.
We provide experimental results based on the evaluation of the proposed scheme over the NITOS Future Internet (FI) facility [17]. NITOS is a heterogeneous testbed, located in the premises of University of Thessaly, in Volos, Greece. The rich heterogeneity of its provided resources allows us to conduct the designated experiments. We employ an LTE base station, along with a UMTS femto-cell and the respective Core Network. We use a testbed node as a WiFi Access Point, located inside the coverage of both LTE and UMTS. NITOS has been upgraded recently with six mmWave WiGig radio units [18], that are reachable by all the testbed nodes. In order to include WiGig in the Multi-RAT technologies, we use a pair of WiGig nodes, that are reachable and interchangeably usable by the UEs involved in the experiment. The overall topology that we use for our experiments is depicted in Fig. 2. In order to port the system model setup over the testbed equipment, we came up with two algorithms for the UEs and the Core Network.

Algorithm 1 Algorithm for UEs
1: if UE is connected to multiRAT then
2: Receive system report \((p_m, C_m, Q_m) \forall RAT\)
3: Calculate the Utility for each TC per RAT
4: UE sends change/OK message to Core Network
5: else
6: if UE decides to connect to multiRAT then
7: Send connect message to Core Network
8: end if
9: end if
10: while 1 do
11: if UE decides to leave State 1 then
12: Send leave message to Core Network
13: else
14: Wait for system report
15: Calculate Utility Function for each TC per RAT
16: UE sends change/OK message to Core Network
17: end if
18: end while

Algorithm 2 Algorithm for Core Network
1: Assign randomly the UEs’ TCs to 4 RATs
2: Calculate prices, available capacities and mass of TCs
3: Create system report \((p_m, C_m, Q_m) \forall RAT\)
4: Create an order of UEs based on (Data Rate Demand/Spectrum Efficiency/Sensitivity)
5: for UEs in system do
6: Send system report to UE
7: Wait for UE response
8: Update system report
9: end for
10: while 1 do
11: Wait update from UEs in State 1
12: Wait for new UEs
13: if Available capacity then
14: Assign new UE’s TCs to RATs
15: else
16: Deny access to new UE
17: end if
18: Calculate prices, available capacities and mass of TCs
19: Create system report \((p_m, C_m, Q_m) \forall RAT\)
20: Create an order of UEs based on (Data Rate Demand/Spectrum Efficiency/Sensitivity)
21: for UEs in system do
22: Send system report to UE
23: Wait for response
24: Update system report
25: end for
26: end while
In the case that a UE enters State 1, the controller determines if the UE can be served by the system and approves or denies access. For every change in the multi-RAT system (UEs entering/leaving State 1), the controller updates the system values and communicates them to the State 1 UEs for further calculation of their utility functions (as shown in Algorithm 1). As seen in Algorithm 2, the Core Network policy for determining the system report depends on the ordering of the UEs based on their data rate demands, spectrum efficiency, and sensitivity. Hence, in the following section we provide results on three different UE ordering policies.

V. System Evaluation

In this section we provide the experimental evaluation of our proposed policy. As the testbed is organized in an RF-isolated setup, we are able to reproduce each experiment. We showcase our experimental results with a resolution of 10 times per experiment, with 6 UEs available in the Multi-RAT system.

Based on the insights from our previous work [4], we evaluate our proposed scheme for three different polling policies.

1) **Policy 1**: The controller is ordering the State 1 UEs based on their data rate demands for all TCs ($\sum_{i=1}^{4} r_i$) in descending order.

2) **Policy 2**: The ordering is based on the normalized spectrum efficiency of the UEs ($w_i$).

3) **Policy 3**: The ordering is based on the UEs’ sensitivity to the changes of the network conditions ($\theta_i$).

In order for MATCH to be evaluated in a real world environment, we determined a setup of clients in the testbed complying with the needs of our model. The rate adaptation algorithms were disabled for all the technologies and the highest Modulation and Coding Scheme was configured for all of them. We measure the capacity of our links by using frames of 1500 bytes and we set the maximum achieved speed for the used technologies to be 42 Mbps for UMTS/HSPA+, 70 Mbps for LTE 10MHz 2x2 MIMO, 130 Mbps for IEEE802.11n 20MHz 2x2 MIMO and 1 Gbps for WiGig. For the configuration of the TCs of each UE, we assume 4 different classes with limits in the lowest and highest throughput that the UEs request from the network. Each UE selects uniformly the throughput value from this range, configured as follows: 1) Background class (0.3-1Mbps), 2) Interactive class (1.1-2.5 Mbps), 3) Streaming class (2.6Mbps - 8Mbps) and 4) Conversational class (8.1-25Mbps). The limits have been set based on [19].

We configure the probabilities of State 0 with $p_{02}$ to be equal to 0.2, representing the probability that UEs on State 0 follow a route outside our system across all our experiments. In addition, the probability of a UE in State 0 to enter the Multi-RAT system ($p_{01}$) is equal to 0.5 and thus the probability of staying in the same state ($p_{00}$) is 0.3. For State 1, the probability a UE exits the system going to State 2 ($p_{12}$) is configured to 0.45 and to remain in the same state ($p_{11}$) to 0.5, meaning that the probability of returning to State 0 ($p_{10}$) is 0.05. For State 2, we set the probability of a UE to remain in this state ($p_{22}$) equal to 0.9 and the probabilities $p_{20}$ and $p_{21}$ equal to 0.05.

<table>
<thead>
<tr>
<th>UE ID</th>
<th>Throughput (Mbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UE0</td>
<td>45.08805063</td>
</tr>
<tr>
<td>UE1</td>
<td>74.65152835</td>
</tr>
<tr>
<td>UE2</td>
<td>110.3614036</td>
</tr>
<tr>
<td>UE3</td>
<td>114.2742001</td>
</tr>
<tr>
<td>UE4</td>
<td>31.71801161</td>
</tr>
</tbody>
</table>

The scenario of the experiments that we present in this section is the following: The experiment starts with UE0 and UE1 in State 1, while UE2, UE3, and UE4 will be entering the system later. UE5 chose the direct route to State 2 and remained there. For the experiment under consideration the aggregate achieved throughput for all traffic classes for each UE in the system is presented in Table I.

Figure 3 is showcasing our averaged results from running the aforementioned use case in the testbed. Figure 3a demonstrates how many times the UEs determined a RAT change for their TCs during the experiment (Desired Switches) versus how many times they accomplished to do so (Executed Switches). As Algorithm 2 recalculates capacities, prices and mass of TCs per each RAT, it may infer that the Utility calculated per each UE for the case of a RAT change may not be valid, and subsequently denies the RAT selection. As it is shown, when the criterion for the ordering is the normalized spectrum efficiency (2nd Policy), the UEs have more freedom to change RATs compared to 1st and 3rd case. In the case of applying the 3rd Policy, we note that the UE ordering at the Core Network renders any RAT selections rather difficult.

Following, we examine how the different policies affect the
charges of the multi-RAT system. In order to compare the three policies, we normalize the costs as shown in Figure 3b. As indicated, the UE with the highest data rate demands will pay more compared to the other clients in the system, at least for the two of the three cases. When applying the 3rd Policy, we observe that the UEs with the highest demands (UE2 and UE3) will pay less than the next UE in order (UE1). Due to this observation, we present in Fig. 3c the normalized cost per Mbps for each UE for each policy. Fig. 3c confirms that the 3rd Policy is not fair in terms of cost for utilizing the multi-RAT. UE4 with the lowest data rate demand will be charged the highest amount per Mbps.

Finally, we seek to investigate the impact of the ordering policy in the utilization of the available RATs of the system. For this purpose, we visualize, in terms of percentages, each UE’s usage of each RAT for each one of the three policies as shown in Figure 4. We monitor that some UEs are affected more than others in the allocation of their TCs to the available RATs, based on the order that they will be probed to determine changes (or not) by the network controller. For instance UE0 and UE3 are following the same pattern through the different ordering policies, with almost the same distribution for the 1st and 3rd policy (Fig. 4a, 4c) and greatly differ for the 2nd (Fig. 4b). On the other hand, UE1 and UE2 are keeping similar distribution behavior across the three policies.

VI. CONCLUSION

In this work we propose MATCH, a dynamic pricing scheme for network selection in a Multi-RAT environment where UEs have multiple traffic classes and are able of concurrently using multiple access technologies. For the evaluation of our scheme, we presented experimental results obtained from the application of our proposed model in a real testbed environment. Through these experiments, we pinpoint the importance of the UEs’ ordering policies (based on data rate demands, spectrum efficiency, and sensitivity) and how they may affect the cost incurred at each UE as well as the utilization of the available RATs. In the near future, we foresee extending our work towards locating the Core Network controller closer to the edge, and incorporating edge services that will support lower latency creating room for new charging policies for the clients.

REFERENCES