

This is a preprint version of the document

Sina Khatibi and Alba Jano, "Elastic Slice-Aware Radio Resource Management with AI-Traffic Prediction", *European Conference on Networks and Communications (EuCNC)*, Valencia, Spain, 2019.

©2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Elastic Slice-Aware Radio Resource Management with AI-Traffic Prediction

Sina Khatibi, Alba Jano Nomor Research GmbH Munich, Germany {khatibi, jano.stud}@nomor.de

Abstract— Network virtualisation and network slicing are the two essential innovations in the next generation of mobile networks also known as the 5G networks. Based on these innovations, multiple network slices with different requirements and objectives can share the same physical infrastructure. The techniques to efficiently allocate the available radio resources to different slices based on their requirements and their priority, also known as inter-slice radio resource management, has been the subject of many studies. The formerly proposed algorithms either assume the slices request maximum contracted data rates or they react passively as the demands arrive. This paper proposes to use Artificial Intelligence (AI) approaches to learn the pattern of the traffic demand of each network slices and predict the demands in the next decision interval. Based on the prediction of the slices' demands, a novel model for elastic inter-slice radio resource management is proposed to increase the multiplexing gain while not compromising the quality of offered connectivity services to the slices. The proposed model is evaluated using a practical scenario. The numeric results show that while the performance of the model under full demand is similar to former models, its elastic resource management enables more efficient resource allocation when the traffic demands vary over time.

I. INTRODUCTION

The emergence of traffic-hungry smart-phones and the massive Internet of Things (mIoT) has hugely influenced the mobile networks. Operators need new solutions to face the rapid growth of mobile traffic demand [1] with a wide range of service Quality of Service (QoS) requirements. Hence, recent studies on the development of the fifth generation of mobile networks (5G) consider flexibility, scalability, and cost efficiency to be the essential design goals and proposed the concept of network virtualisation. The aim is to modernise the mobile network by changing how operators architect their networks using Software Defined Network (SDN) and Network Function Virtualisation (NFV) architectures [2].

The key idea in both SDN and NFV is network element abstraction, based on which different functionalities of nodes in a system can be implemented either as SDN applications or Virtual Network Functions (VNFs). These two architectural concepts propose to separate the physical network infrastructure from the nodes' functionalities. This separation enables the implementation of network functions on Commercial Off-The-Shelf (COTS) instead of proprietary hardware [3]. The network virtualisation also allows the sharing of infrastructure among multiple network slices while offering isolation. The operators, hence, can share server farms, over which they deploy VNF chains to form their networks and reduce their CAPital EXpenditure (CAPEX)/Operational EXpenditure (OPEX) while increasing the flexibility of their systems.

Also, these innovations are going to change business models and stakeholder models. One of these changes is the emergence of Small and Medium Enterprises (SMEs) as VNF providers to offer more diversity to the market, which is currently ruled by giant well-established mobile vendors. Hence, operators have more options to choose, and they can establish a multi-vendor network, which fits the best to their needs and business strategies.

However, the realisation of a full-fledged virtual mobile network requires addressing essential challenges such as slice-aware resource management. Serving multiple network slices with different QoS requirements over the same physical infrastructure is a non-trivial task [4]. Each network slice requires a portion of available radio resources and sufficient computational resources to run the related VNF chains. The tight latency requirements in RAN imposes rigid design constraints on the processing delay budget, provisioning and allocation of computational resources. This task becomes even more laborious because different VNFs from different providers would have different computational performance.

The formerly proposed inter-slice Radio Resource Management (RRM) algorithms either allocate resources in passive mode (i.e. adopt the allocation after observation of demand change) [5] or they allocate based on maximum demand assumption [6, 7]. Dynamic reservation of resources for the slices is also another approach to inter-slice radio resource management [8]. In [9], authors proposed a Q-learning approach to autonomous adjust the resource reservation. However, to enable elastic resource allocation to improve radio resource utilisation and increase the multiplexing gain, all these algorithms require to predict the slice's traffic demand. This prediction is even more critical in computational resource orchestration and management since creating a new instance over virtual machines in the cloud environment can take up to several hundred milliseconds. Provisioning of the VNF chains for the highload leads to extra CAPEX and OPEX, on the one hand. On the other hand, the low-load provisioning may leads to computational resource shortage and service intruption.

The novelty of this paper is to propose a new model for elastic slice-aware radio resource management. The novel model proposes to add Deep Neural Networks (DNNs) to resource manager algorithm to predict the slices demands and improve resource provisioning in mid-term (i.e. in the order of a couple of minutes to a couple of hours). The algorithm with a realistic estimation of future demands can

Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

elastically manage the resources and improve the quality of service offered to slices. Also, the elastic resource management improve the resource utilisation, and it leads to the cost reduction.

The rest of this paper is organised as follows: Section II describes the slice traffic prediction, and Section III addresses the new elastic radio resource management. Section IV provides the scenario description and numerical results. Finally, Section V concludes the paper.

II. PREDICTION OF SLICES' DEMANDS

Depending on the services of a slice, their traffic demands during the time interval as long as a day or week follows a similar pattern. It is common to observe a gradual increase of demands up to the rush hour (i.e. the busiest time interval of the day) followed by a slow decay till the next day. The demands pattern not only varies temporarily but also geographically. The commercial zone experiences higher demands during the working hours while the residential areas experience higher load in the afternoon.

The paper proposes to train DNNs with the slices' traffic demand in a serving area based on long-term reports to predict the demand based on the N-previous observational intervals. Retraining of the neural network by the new reports (e.g. after a day or each couple of hours) enables it to cope with slow and long-term changes in the traffic pattern. Fig. 1 illustrates the DNN for demand predictor in this paper, which has 5-neuron input layer, followed by two 16-neuron dense layers with ReLU activation function in addition to a sigmoid activation function for the output layer [10].

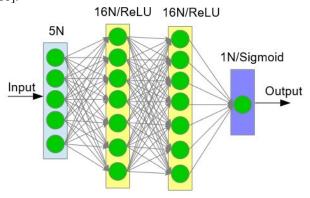


Fig. 1. The structure of the DNN.

Fig. 2 shows the predicted against the actual traffic. It is apparent from the graphs that the predictions are following the average of the traffic time in each interval. The interslice model can use these predictions as the input to elastically allocate resources to each slice. It is worth noting that the neural network cannot predict the sudden variation of traffic, but its prediction is closer to the average demand in the next time interval.

III. ELASTIC SLICE-AWARE RADIO RESOURCE MANAGEMENT MODEL

The allocation of available radio resources to different network slices can be formulated as an optimisation problem to maximise the weighted network throughput subject to constraints including the Service Level Agreements (SLAs), the total network throughput, and fairness in addition to underprovisioning and violation penalties.

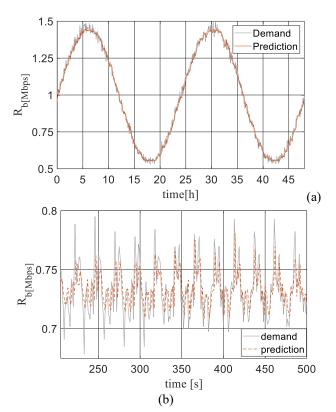


Fig. 2. Traffic demand prediction using the deep neural network.

The proposed model in this paper is the extension of the model described in [7] to consider the traffic demand of each slice. The former models [7, 11, 12] assumed that the network slices always have higher demands than the allocated radio resources; Hence, they can always fully utilise the radio resources allocated to them. In the new model, the deep neural networks first predict the traffic demands and this prediction can be corrected with a confidence coefficient:

$$R_{b_{i[\text{Mbps}]}}^{Pre} = \gamma_i R_{b_{i[\text{Mbps}]}}^{\widetilde{Pre}}$$
(1)

where:

- $\frac{R_{b_i}^{Pre}}{R_{b_i}^{Pre}}$: the confidence demand prediction for slice *i*, $\overline{R_{b_i}^{Pre}}$: the demand predicted by the DNN for slice *i*, γ_i : the confidence factor for slice *i*, $\gamma_i \ge 0$.

In addition to traffic prediction, the inter-slice radio resource management model has two other essential parts: i) Estimation of the total network throughput, ii) Allocation of available resource to the network slices.

Based on the probability functions of Signal to Noise-Interference (SINR), the probability functions of the throughput of a single Physical Radio Block (PRB) is calculated. The total network throughput is the convolution of all PRBs probability density function as described in [13].

The model for allocation of available radio resources to different network slices is a multi-objective optimisation model. These objectives while related, they may be contradictive. The objectives are as follows:

Slice Priority is the primary objective in the proposed model and aims to allocate more radio resources and higher throughput to the slices with higher priority. Hence, the related objective function in the model is the weighted throughput given by:

$$f^{Srv}(\boldsymbol{R}_{\boldsymbol{b}}^{Srv}) = \sum_{i=0}^{N_S} w_i^{Srv} R_{b_{i[\text{Mbps}]}}^{Srv}$$
(2)

where:

- f^{Srv} : slice priority objective function,
- R_b^{Srv} : vector of serving throughputs, R_b^{Srv} : serving (allocated) throughput to slice *i*, •
- N_s : number of network slices, w_i^{Srv} : serving weight of slice *i*.

The serving weights in the equation above define the priority of slices. The slice with relatively higher serving weight has relatively higher priority, and the model tends to allocate more resources to it comparing to the other slices. It is common practice to have the summation of the serving weights equal to unit.

Understocking is the focus of the second objective function. It is ideal to allocate to each slice what they are going to demand. However, there are cases where there are not enough resources available to serve all the requested throughput from all the network slices. In these situations, the allocated throughput to each slice is going to be smaller than the predicted demand. The understocking, in the framework of this paper, is a non-negative value equal to the difference of the predicted demand of a slice and the serving throughput of the same slice. The model tries to minimise the summation of understocking throughput given by:

$$f^{us}(\boldsymbol{R}_{\boldsymbol{b}}^{\boldsymbol{us}}) = \sum_{i=0}^{N_s} R_{b_{i[\text{Mbps}]}}^{us}$$
(3)

where:

- f^{us} : the underprovioninb objective function, ٠
- ٠
- R_b^{us} : vector of underprovioninb throughputs, R_{bi}^{us} : understocking throughput to slice *i*, given by: •

$$R_{b_i}^{us} = \min\left\{0, R_{b_{i[\text{Mbps}]}}^{Pre} - R_{b_{i[\text{Mbps}]}}^{Srv}\right\}$$
(4)

It is worth noting that the confidence factor introduced in (1) contains the overstocking (i.e. allocation of resources to a network slice more than the predicted demand.

Violation of Service Level Agreements (SLAs) is the next objective in the inter-slice radio resource management model. The allocated resources to each network slice should meet the requirements dictated by the SLA of the slice. While slices may have different SLAs, as introduced in [7], the three main SLAs categories are as follows:

Guaranteed Bitrate (GB): This type of SLA 1. guarantees the network slice throughput to be higher than the minimum guaranteed data rate and less than maximum guaranteed data rate, given by

$$R_{b_{i[\text{Mbps}]}}^{min} \le R_{b_{i[\text{Mbps}]}}^{Srv} \le R_{b_{i[\text{Mbps}]}}^{max}$$
(5)

where:

- $R_{b_i}^{min}$: minimum guaranteed data rate, $R_{b_i}^{max}$: maximum guaranteed data rate.
- Best effort with minimum Guaranteed (BG): 2. The minimum throughput for the network slice is guaranteed in this type of SLA. However, the allocation of more resources is subject to the availability of resources.

$$R_{b_{i[\text{Mbps}]}}^{min} \le R_{b_{i[\text{Mbps}]}}^{Srv} \tag{6}$$

3. Best Effort (BE): A network slice with the best effort SLA is only served when there are resources available, and no throughput is guaranteed.

So, the model should not allow any violation of SLAs. Under circumstances where the violation of SLAs is unavoidable (e.g. when there are not enough resources) then the summation of violations has to be minimised. The SLAs violations, R_{b}^{v} , is the difference between the demanded guaranteed data rate, given by:

$$R_{b_{i[Mbps]}}^{v} = \min\left\{R_{b_{i[Mbps]}}^{Pre}, R_{b_{i[Mbps]}}^{min}\right\} - R_{b_{i[Mbps]}}^{Srv}$$
(7)

Fairness is the final objective function of the new model for inter-slice radio resource management. Although it is desirable to serve the slice with relatively higher serving weights (i.e. high priority slices) with higher throughput, not serving the other slices is not an acceptable output. The final objective in the model ensures that the slices receive the predicted demand data rate when there are enough resources. However, the fairness objective in this model makes sure the violation of SLAs and understocking are distributed among all the slices based on their priority in the congestion situations.

In the case of understocking, demands and serving weight are the two parameters, which the fairness considers. Regarding demands, it is only fair if the slices with high demands experienced higher. For example, 10 Mbps understocking for a slice with 100 Mbps demand as well as a slice with 10 Mbps is not fair. In the former case, the slice received 90% of what it has demands while in the latter case the slice is not served all. Now assume that both slices have the same amount of demand, but their serving weights are different. The fair allocation suggests the slice with higher serving weight (i.e. the slice with higher priority) experience less understocking.

Hence, a fair resource allocation considering understocking is when the condition below is achieved:

$$\forall i: \left| \frac{w_i^{Srv} R_{b_{i[Mbps]}}^{us}}{R_{b_{i[Mbps]}}^{pre}} - \frac{1}{N_s} \sum_{j=0}^{N_s} \frac{w_j^{Srv} R_{b_{j[Mbps]}}^{us}}{R_{b_{j[Mbps]}}^{pre}} \right| = 0$$
(8)

In the case of violations, the fairness only considers the violation weights, w_i^{ν} . Thus, the fairness condition in this situation is:

$$\forall i: \left| w_i^{\nu} R_{b_{i[Mbps]}}^{\nu} - \frac{1}{N_s} \sum_{j=0}^{N_s} w_j^{\nu} R_{b_{j[Mbps]}}^{\nu} \right| = 0 \tag{9}$$

However, it is always neither possible (i.e. the constraint imposed by SLA) nor desirable to satisfy the conditions stated in (8) and (9) but the goal is to minimise the diversion from these conditions. Hence, the objective function for fairness is:

$$f^{f}(\boldsymbol{R}_{b}^{f_{us}}, \boldsymbol{R}_{b}^{f_{v}}) = \sum_{j=0}^{N_{s}} R_{b_{j[Mbps]}}^{f_{v}} + R_{b_{j[Mbps]}}^{f_{us}}$$
(10)

where:

- f^{f} : fairness objective function, $R_{b}^{f_{us}}$: vector of understocking fairness derivation, $R_{b_{i}}^{f_{us}}$: understocking fairness derivation for slice *i*,
- $R_{b}^{f_{v}}$: vector of violation fairness derivation,
- $R_{b_i}^{f_v}$: violation fairness derivation for slice *i*,

Based on the objectives discussed above, a multiobjective linear program optimisation problem as follows:

£

. . .

$$\max_{R_{b}} f^{Srv}(\mathbf{R}_{b}^{Srv}) - \alpha_{f} f^{f}(R_{b}^{f_{v}}, R_{b}^{f_{us}}) + \alpha_{v} f^{v}(R_{b}^{v}) + \alpha_{us} f^{us}(R_{b}^{us})$$

$$= \left\{ \begin{vmatrix} \sum_{i=0}^{N_{s}} R_{b_{i[Mbps]}}^{Srv} \leq R_{b[Mbps]}^{tot} \\ \frac{w_{i}^{Srv} R_{b_{i[Mbps]}}^{us}}{R_{b_{i[Mbps]}}^{pre}} - \frac{1}{N_{s}} \sum_{j=0}^{N_{s}} \frac{w_{j}^{Srv} R_{b_{j[Mbps]}}^{us}}{R_{b_{j[Mbps]}}^{pre}} \end{vmatrix} \leq R_{b_{i}[Mbps]}^{f_{srv}}$$

$$= \left\{ \begin{vmatrix} w_{i}^{v} R_{b_{i[Mbps]}}^{v} - \frac{1}{N_{s}} \sum_{j=0}^{N_{s}} w_{j}^{srv} R_{b_{j[Mbps]}}^{us}} \\ w_{i}^{v} R_{b_{i[Mbps]}}^{v} - \frac{1}{N_{s}} \sum_{j=0}^{N_{s}} w_{j}^{v} R_{b_{j[Mbps]}}^{v} \end{vmatrix} \leq R_{b_{i}[Mbps]}^{f_{v}} \\ \min \left\{ R_{b_{i}[Mbps]}^{Pre}, R_{b_{i}[Mbps]}^{min} \right\} - R_{b_{i}[Mbps]}^{Srv} \leq R_{b_{i}[Mbps]}^{v} \\ R_{b_{i}[Mbps]}^{Srv} \leq R_{b_{i}[Mbps]}^{max} \end{cases}$$

where:

- R_{h}^{tot} : the total network throughput,
- α_f : weight of fairness objective,
- α_{ν} : weight of voilation objective,
- α_{us} : weight of understocking objective.

It worth noting that the desired values for the fairness, violation, and understocking objective function are zero while as the total network throughput increases, the weighted sum of the slice data rate, i.e. the main objective function, increases. Hence, the weights of the other objective functions should be relative to the total network throughput. However, this paper considers comparative big weights (>10000 times of the total network's throughput) as the objective weights.

IV. NUMERICAL RESULTS

The chosen scenario to evaluate the performance of the inter-slice radio resource management model is based on the scenarios described in [12]. The scenario considers a serving area, which contains 16 cells with a radius of 400 m each with 500 PRBs. The total network throughput estimated as described in Section III is 1.2 Gbps.

The terminals in this scenario require the average throughput of 13 Mbps for each terminal [1] and the network slices contract data rates relative to their number of terminal.

$$R_{b[\text{Mbps}]}^{Cont} = N_{ue} \overline{R_{b[\text{Mbps}]}^{ue}}$$
(12)

where:

- R_b^{Cont} : the contracted data rate of slice, N_{ue} : the number of User Equipment (UE), $\overline{R_b^{ue}}$: the average data rate per UE.

Also, there are three network slices with three different SLAs. Slice GB has guaranteed data rate SLA, and the allocation of resources has to guaranteed throughput between 50% up to 100% of its contracted data rates. The BE slice has a minimum guaranteed data rate of 25% of its contracted data rate. Finally, the allocation of resources to slice BE is subject to availability or resources. All three slices have the same services and the same contracting data rate. Table I summaries the serving, understocking, and violation weight for different slices.

In the first step, the performance of the inter-slice radio resource management proposed in this paper is evaluated under the full-demand mode (i.e. when the network slices

demand all the contracted data rates). In this case study, the number of UEs per slice is swept from 10 active terminals up to 350 terminals. Fig. 2 illustrates the allocated throughput to each network slices as the function of the number of UEs per slice.

TABLE I. WEIGHTS AND THE SLAS FOR THE SLICES.

	W_i^{Srv}	w_i^v	$R^{min}_{b_{i[extsf{Mbps}]}}$	$R_{b_{i[Mbps]}}^{max}$
GB	0.07	0.63	$0.5R_b^{Cont}$	R_b^{Cont}
BG	0.02	0.18	$0.25R_b^{Cont}$	N/A
BE	0.01	0.09	0	N/A

According to the figure, the total traffic demands per slice increases up to 1.5 Gbps. Given the 1.2 Gbps network capacity, it is not possible to serve all the slices when all off have the full demanded throughput. The graph shows that the understocking (i.e. not serving the demanded throughput) starts happening when the demand per slice reaches to 0.38 Gbps. As the demands per slice increases, slice GB with the highest serving weights comparatively receives higher throughput than the slice BE with lowest serving weights. The slice BE does not receive any more resources. The violation to the minimum guaranteed data rate of slice BG happens when the total number of UEs reach to 80 terminals and they demand the total throughput of 1.05 Gbps.

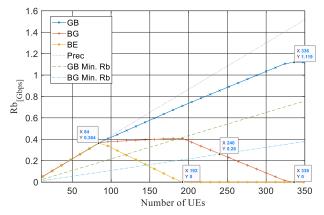


Fig. 3. The allocated throughput to each network slice in full demand.

In the next step, the number of UEs per slice is fixed to 180 but they are not demanding all the contracted traffic. The traffic of slice BG is fixed on 45% of contracted data rates equal to 2.05 Gbps while the demand of the slice GB is increasing the demand for slice BE is decreasing. Fig. 3 presents the predicted demand (shown by "Prec" prefix) versus the allocated throughput in 20 observation intervals.

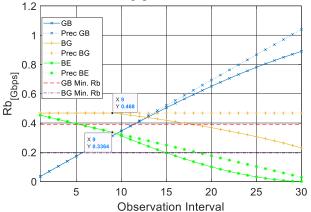


Fig. 4. Resource allocation to slices with demand prediction.

It is apparent that the slice BE and BG have also received the full demanded throughput up to ninth observation interval, where the demand of the slice GB passed 336.4 Mbps — comparing the numeric results from Fig. 2 and Fig. 3 demonstrates the multiplexing gain using the throughput prediction. Using the former algorithms, the slice BE would not have received any throughput if the changes of other slices' were not taken into account. The slice BG also is experiencing higher data rates since the most important slice, the slice GB, is not fully requesting its contracted data rates.

It can be concluded through the second scenario, the new model increases the elasticity of resource allocations and the better quality of service to slices using the same physical infrastructure.

V. CONCLUSIONS

The network function virtualisation is an essential innovation in the 5G systems. It enables serving multiple network slices over the same physical cloud-based infrastructure to reduce CAPEX and OPEX and increase the flexibility of the system.

Inter-slice radio resource management algorithm is in charge to allocate the available radio resources from the infrastructure resource pool to the slices. Serving different network slices with different SLAs and objective is a challenging task. The formerly proposed algorithms do not consider the variation of slices' traffic demands and always assign them the maximum possible resources based on their serving priority and their SLAs.

However, a slice-aware elastic radio resource management algorithm can improve the flexibility of network even more and avoid over-allocation of resources to the slices when they may not demand it. Hence, the network can allocate more resources to the slices with lower serving priority but higher traffic demands. It also enables the infrastructure providers to admit more network slices with the same resource pool and reduces their costs.

This paper proposed a novel AI-based elastic sliceaware radio resource management algorithm, which uses deep neural networks to learn the pattern of the slices traffic demand and use it to predict the future slices' demand. The main design assumption for the neural network is based on the fact that the traffic demands in the networks change gradually. Hence, the expected traffic volume can be estimated by the changes in traffic during the last Nobservation intervals.

The resource allocation model is also extended to allocate each slice only the required radio resources and avoid unnecessary overstocking. The model considers the serving weights and SLAs of the slices and allocates more resources to slice with higher priority. When there are not enough resources available, the model aims to minimise the violations to the slices' SLAs and the understocking. The fairness objective functions ensure that both understocking and violations are fairly distributed among the slices.

The performance of the proposed model is also evaluated through a series of practical scenarios. The numerical results suggest that the model allocates more resources to slices with higher priority when they are demanding their full contracted data rates. However, the results clearly present the elastic resource allocation when there are changes in the slice traffic demands. Comparing the allocated resources to the slices, it is apparent that there is a reduction of allocation to the slices resources when they have lower traffic demands. As the demands increases, the share of the slice from the resource pool relative increases. In the case study, the throughput of slice BE with the lowest serving priority has increased up to 45.34 Mbps.

ACKNOWLEDGEMENT

Part of this work has been performed within the 5GMoNArch project, part of Phase II of the 5th Generation Public Private Partnership (5G-PPP) program partially funded by the European Commission within the Horizon 2020 Framework Program.

REFERENCES

- Cisco Systems, "Cisco Visual Networking Index: Forcast and Methodology, 2017-2022," in "from Visual Network Index (VNI) White Paper," Cisco Systems, CA, USA, Nov. 2018.
- [2] Y. Li and M. Chen, "Software-Defined Network Function Virtualization: A Survey," *IEEE Access*, vol. 3, pp. 2542-2553, Dec. 2015.
- [3] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, "Network Slicing and Softwarization: A Survey on Principles, Enabling Technologies, and Solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2429-2453, Mar. 2018.
- [4] S. Khatibi, I. Balan, and D. Tsolkas, "Slice-Aware Elastic Resource Management," presented at the EuCNC'18 -- IEEE European Conference on Networks and Communications - Special Session Resource Elasticity for 5G Network Architecture, Ljubljana, Slovenia, 2018.
- [5] L. Caeiro, F. D. Cardoso, and L. M. Correia, "Wireless access virtualisation: Addressing virtual resources with different types of requirements," in *European Conference on Networks and Communications (EuCNC)*, Bologna, Italy, 2014, pp. 1-5.
- [6] B. Rouzbehani, L. M. Correia, and L. Caeiro, "A Service-Oriented Approach for Radio Resource Management in Virtual RANs " *Wireless Communications and Mobile Computing*, vol. 2018, p. 13, 2018, Art no. 4163612.
- [7] S. Khatibi and L. M. Correia, "A model for virtual radio resource management in virtual RANs," *EURASIP Journal on Wireless Communications and Networking*, journal article vol. 2015, no. 1, p. 68, Mar. 2015.
- [8] L. Caeiro, F. D. Cardoso, and L. M. Correia, "Wireless access virtualisation: Addressing virtual resources with different types of requirements," in 23rd IEEE European Conference on Networks and Communications (EuCNC'14), Bologna, Italy, 2014, pp. 1-5.
- [9] G. Sun, Z. T. Gebrekidan, G. O. Boateng, D. Ayepah-Mensah, and W. Jiang, "Dynamic Reservation and Deep Reinforcement Learning Based Autonomous Resource Slicing for Virtualized Radio Access Networks," *IEEE Access*, vol. 7, pp. 45758-45772, Apr. 2019.
- [10] D. M. Gutierrez-Estevez et al., "Artificial Intelligence for Elastic Management and Orchestration of 5G Networks," *Submitted to IEEE Communication Magazine*, 2019.
- [11] B. Rouzbehani, L. M. Correia, and L. Caeiro, "A modified proportional fair radio resource management scheme in virtual RANs," in 2017 European Conference on Networks and Communications (EuCNC), 2017, pp. 1-5.
- [12] B. Rouzbehani, L. M. Correia, and L. Caeiro, "An Optimised RRM Approach with Multi-Tenant Performance Isolation in Virtual RANs," in *IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'18)*, Bologna, Italy, 2018, pp. 1-6.
- [13] S. Khatibi and L. M. Correia, "The Effect of Channel Quality on Virtual Radio Resource Management," in *IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, Boston, MA, USA, 2015, pp. 1-5.