Load-Aware Cell Switching in Ultra-Dense Networks: An Artificial Neural Network Approach

Attai Ibrahim Abubakar*, Metin Ozturk*, Rao Naveed Bin Rais†, Sajjad Hussain*, Muhammad Ali Imran*

***Abstract*—Most online cell switching solutions are sub-optimal because they are computationally demanding, and thus adapt slowly to a dynamically changing network environments, leading to quality-of-service (QoS) degradation. This makes such solutions impractical for ultra-dense networks (UDN) where the number of base stations (BS) deployed is very large. In this paper, an artificial neural network (ANN) based cell switching solution is developed to learn the optimal switching strategy of BSs in order to minimize the total power consumption of a UDN. The proposed model is first trained offline, after which the trained model is plugged into the network for real-time decision making. Simulation results reveal that the performance of the proposed solution is very close to the optimal solution in terms of trade-off between the power consumption and QoS.**

I. Introduction

Base station (BS) switching is one of the most generally accepted techniques for energy saving in mobile cellular networks (MCNs) [1]. It takes advantage of the spatio-temporal variations in user traffic demands to match power consumption with traffic demand per time thereby avoiding energy wastage during period of low or no traffic load. Various BS switching schemes have been proposed in literature employing analytical [2], heuristic [3], and machine learning [4], [5] techniques.

Among them, machine learning implementations are beginning to gain more application in wireless communications because of their ability to learn complex network behaviours that are hard to be accurately modelled analytically. They also have the ability to adapt to dynamic network environment, which is difficult for most hard-coded heuristic algorithms [6]. In addition, machine learning models can be trained offline and then plugged into the network in order to enhance real-time decision making by reducing the network delays [7]. One of such machine learning algorithms is the artificial neural networks (ANN), which are known as universal approximators because they are able to find the relationships between complex non-linear functions and are also known for their excellent generalization ability [8], hence they have found numerous applications in the field of wireless communications [7].

A few research works have been carried out regarding the application of ANN for cell switching purposes [4], [9], [10]. In [4], an ANN algorithm was proposed to determine the switching pattern that maximizes the energy efficiency of the network while ensuring that the minimum bit rate requirements of the users is satisfied. A context-based energy saving approach for cache enabled BSs using Bayesian neural networks was proposed in [9]. The authors in [10] applied ANN for traffic prediction and cell switching decisions with two different ANN architectures.

However, the previously mentioned solutions only consider simplistic network scenarios where very few small cells (SCs) are deployed, hence, such solutions will not be suitable when network dimensions becomes very large and complexity increases. In addition, only one type of SC was considered in the aforementioned works which is not the case in a real network where different types of SCs (remote radio head (RRH), micro, pico, and femto cell) are deployed, thus making their considered scenarios unrealistic.

In this paper, we exploit ANN to determine the optimal cell switching strategy in an ultra-dense network (UDN). Specifically, we develop an ANN-based cell switching framework, which is referred to as offline-trained online cell switching (OTOcell), to learn the optimal switching strategy for the SCs in a UDN. The developed model is computationally efficient since the training is done offline, after which the trained model is implemented in the network for real-time decision making. This is particularly important for UDNs, where the macro cells (MCs) are already over-burdened with signalling and computational operations, in which case, adding a cell switching algorithm on top of these would make their workload more severe. We also consider various types and number of SCs to validate the robustness and scalability of the proposed solution.

The remaining parts of the paper is organized as follows: Section II presents the system model, followed by a description of the proposed OTOcell framework in Section III. Section IV evaluates the performance of the model while Section V concludes the paper.

II. System model

A. Network model

A heterogeneous UDN, with separate control and data plane, comprising both MC and SCs is considered. Four types of SCs (RRH, micro, pico, and femto cell) are considered in the work. The MC—which encompasses the footprints of the SCs—is constantly kept on, and also orchestrates the switching operation of the SCs via its backhaul connection to them. The SCs, on the other hand, can be turned ON/OFF based on their traffic load and are responsible for handling high data traffic demands. Vertical traffic offloading is considered, such that the traffic load of the SCs that are switched OFF are offloaded to the MC to ensure that the quality-of-service (QoS) of the offloaded users are maintained.

*James Watt School of Engineering, University of Glasgow, UK.
†Electrical and Computer Engineering, Ajman University, UAE.

B. Power Consumption model

The power consumption model of a BS proposed in [11] is adopted and is expressed as:

$$P_{\text{BS}} = \begin{cases} P_{\text{c}} + \sigma_y P_{\text{tx}}, & \text{if } 0 < P_{\text{tx}} < P_{\text{m}} \\ P_{\text{s}}, & \text{if } P_{\text{tx}} = 0, \end{cases} \tag{1}$$

where P_{BS} represents the BS total power consumption, σ_{y} denotes the slope of the load dependent components, P_{c} is the constant power consumption component of the BS, and P_{s} is the sleep mode power consumption. P_{tx} and P_{m} denote the instantaneous and maximum BS transmission power, respectively. The relationship between the traffic load of the BS and transmission power can be expressed as:

$$P_{\text{tx}} = \tau P_{\text{m}}, \qquad 0 \leq \tau \leq 1 \tag{2}$$

where τ is the normalized traffic load of the BS.

The total power consumption of the UDN comprises the power consumption of the MC and that of all the SCs deployed under its coverage. This can be expressed as:

$$P_{\text{T}} = P_{\text{MC}} + \sum_{n=1}^{N} P_{\text{SC},n}, \tag{3}$$

where P_{T}, P_{mc} and $P_{\text{sc},n}$ are UDN's total, MC's, and n-th SC's power consumption, respectively.

C. Problem Formulation

The aim of this research is to select the optimal combination of SCs to switch OFF, during periods of low or no traffic in order to minimize the total power consumption of a UDN while ensuring that the QoS of the users originally connected to the switched OFF SCs are maintained by the MC.

Hence, the optimization objective can be defined as:

$$\begin{aligned} \min_{\phi \in \Phi} \quad & P_{\text{T}}(\phi) \\ \text{s.t.} \quad & \hat{\tau}_{\text{MC}} \leq 1. \end{aligned} \tag{4}$$

where ϕ is the selected SC switching policy, while Φ is the set of all the possible SC switching combinations. $P_{\text{T}}(\phi)$ is the expected power consumption of the network with ϕ switching policy. $\hat{\tau}_{\text{MC}}$ is the traffic load of MC after the offloading is complete, and is given as

$$\hat{\tau}_{\text{MC}} = \tau_{\text{MC}} + \sum_{n=1}^{N} \tau_{\text{SC},n} \Gamma_n, \tag{5}$$

where τ_{MC} and $\tau_{\text{SC},n}$ are the original traffic demands (i.e., before offloading) of MC and n-th SC, respectively, and N is the total number of SCs in the network. Γ is a control parameter, which is responsible for offloading the traffic load of only the switched OFF SCs, such that

$$\Gamma_n = \begin{cases} 1, & \text{if SC}_n \textit{ is OFF} \\ 0, & \text{if SC}_n \textit{ is ON}, \end{cases} \tag{6}$$

where SC_n is the n-th SC.

The constraint in (4) is to ensure that there must be sufficient capacity in the MC to accommodate both the original traffic demand of the MC, τ_{MC}, and the total traffic demand of all the SCs that are switched OFF in order to maintain the QoS.

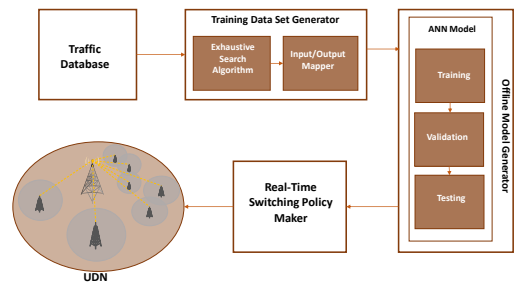

Fig. 1. Overview of the proposed OTOcell framework.

III. OTOcell Framework

Most of the cell switching solutions developed using heuristic approaches, such as exhaustive search (ES) or genetic algorithm, are not suitable for real-time implementation, particularly in networks with large dimensions because they are usually computationally demanding. As a result, before these algorithms decide which set of SCs to switch ON/OFF and execute the decision, the network state would have changed, thereby leading to sub-optimal switching decision and delays.

However, these heuristic approaches can be combined with ANN to accelerate the computation of the optimal cell switching strategy. Our proposed framework is built upon two basic observations: 1) cell switching can be considered to be a problem of deciding the mapping between the traffic demand and optimal switching pattern; 2) ANN are popularly referred as universal function approximators, implying that they can learn the mapping between almost any input and output [8].

Based on these observations, we propose an OTOcell framework to determine the optimal switching strategy that maps the traffic demand of the BSs to the optimal cell switching pattern. The proposed OTOcell framework is summarized in Fig. 1. The traffic loads of the MC and all the SCs associated with it are collected and stored in the *Traffic Load Database*. The traffic loads are then passed to the *Training Data Set Generator* which consists of the ES algorithm, and Input/Output Mapper. The ES algorithm uses the optimization function in (4) to decide the optimal set of SCs to switch ON/OFF per time while the Input/Output Mapper prepares the training data set—which includes the traffic loads, and optimal switching combinations. The training data set is then transferred to the *Offline Model Generator* for ANN model training, validation and testing.

The ANN model utilized is a feed-forward architecture comprising one input layer, three hidden layers (HLs), and one output layer (OL). The number of neurons in the input layer is determined by the input features of the training data set (number of SCs and MC), the number of neurons in the HL are selected empirically by trying different combinations, and the OL neurons is given by 2^N, where N is the number of SCs. The activation function for the HL neurons is ReLU while that of the output neurons is softmax (since the cell switching problem is a multi-classification problem).

TABLE I
PARAMETERS FOR THE DEVELOPED ANN MODEL

Parameter	Scenario 1	Scenario 2
HLs, Neuron size	3, 128 × 128 × 128	3, 32 × 32 × 32
OL neuron size	16	4096
Learning rate	0.0001	0.001
Batch size	30	50
Epochs	1000	1000
HL Activation Function	ReLU	
OL Activation Function	Sofmax	
Loss function	Categorical-crossentropy	
Optimizer	Adam	

The function of the ANN is to learn the mapping between the SC traffic demands and the optimal switching pattern through training. The training process involves adjusting the ANN parameters, using gradient-descent algorithms, such that the difference (error) between the expected output (i.e., predicted output) and the actual output (label) is as close to zero as possible. This error is usually estimated using a loss function and in our case the categorical cross-entropy function is employed. The trained cell switching model is then transferred to the *Real-time Switching Policy Maker* for real-time SC switching.

The justification for using the proposed framework is twofold: 1) once the ANN model is fully trained, the optimal cell switching pattern can be obtained in real-time, that is, whenever the network status changes, without resorting to computing the optimization objective afresh; 2) both training data set generation and the ANN training stage, which are the computationally intensive processes, can be done offline, thus enabling the trained model to be implemented for real-time cell switching without additional computational overhead and delays to the network.

IV. PERFORMANCE EVALUATION

A. Simulation Scenario and data-set generation

Two simulation scenarios, Scenario-A and Scenario-B, with different number of SCs are considered to test the performance of the proposed model on varying network sizes. Both scenarios consists of 1 MC, but Scenario-A has 4 SCs (1 of each type of SC), while Scenario-B has 12 SCs (2 RRH, 3 micro, 4 pico, and 3 femto cells)[1]. The traffic load of both MC and SCs are generated using uniform random distribution model, such that $\tau_{\text{MC}} \in [0, m_{\text{MC}}]$ and $\tau_{\text{SC}} \in [0, m_{\text{SC}}]$ where $m_{\text{MC}}, m_{\text{SC}}$ are the maximum normalized loads of the MC and SCs, respectively. In Scenario-A, BS switching pattern were generated for 7 days with one-minute resolution using ES amounting to about 10,000 observations, while Scenario-B was for 35 days[2] resulting in about 50,000 observations. For the remaining simulation parameters, we adopted values in [11].

[1]The number of each type of SC was selected randomly.

[2]More data set is generated in Scenario-B because the increase in network dimension and complexity makes the training process more difficult.

B. ANN Training and Testing

For Scenario-A, two data sets—each comprising about 10,000 traffic load samples of the BSs and their corresponding optimal switching patterns—were used for training and testing the proposed model. For Scenario-B, one data set comprising about 50,000 traffic load samples and their corresponding optimal switching patterns was utilized, out of which 80% was used for training and 20% for testing. The training of the model in both scenarios is carried out using the Adam optimization algorithm [12]. Table I summarizes the parameters of both models. Upon successful training of both models in each scenario, the trained models are then applied to the test data set in order to evaluate the performance of the trained models.

C. Results and discussion

Figs. 2 and 3 presents a comparison of the total power consumption of the UDN, for both scenarios of OTOcell versus two benchmark approaches: 1) All-ON, which is the conventional approach where no switching is implemented, that is, all the SCs and MC are constantly kept on; and 2) ES approach, which tries to find the best switching policy by considering all the possible switching combinations and selecting the one that results in the least power consumption while the constraint in (4) is satisfied. The ES approach is guaranteed to always return the optimal policy, and hence the goal of any switching technique is to produce the closest approximation of this approach.

In Fig. 2, it can be observed that the performance of the proposed OTOcell is the same as that of ES most of the time but shows slight variations at some time instances due to wrong cell switching prediction from the OTOcell. Compared to the All-ON, it can be observed that the OTOcell shows a reduction in power consumption, however, we notice that due to the few number of SCs, the reduction in power consumption is not significant most of the time as the SCs have fewer opportunities to sleep.

In Fig. 3, where the number of SCs is increased from 4 to 12, the OTOcell shows a slightly lesser performance compared to that of Fig. 2 as the deviation from the optimal ES is more pronounced. This can be traced to the fact that the network dimensions and complexity is increased in Scenario-B compared to Scenario-A, and as a result the OTOcell is prone to more prediction errors. However, compared to the All-ON method, the proposed method shows a significant reduction in power consumption at all time instances owing to the fact the number of SCs has tripled, hence there are more opportunities to switch OFF more SCs.

The QoS evaluation of the proposed framework is also carried out using the throughput metric to ascertain the impact of the OTOcell framework on QoS of the network. Here, the network throughput is considered to be the traffic demand that is supported by all the active BSs in a given time interval. Table II presents the throughput per time slot of the UDN using the OTOcell framework and the two benchmarks. The throughput per time slot is the total network throughput for a given time slot, i.e., every hour and TP avg. is the average

network throughput. It should be noted that not all the time slots are shown in Table II for conciseness purpose.

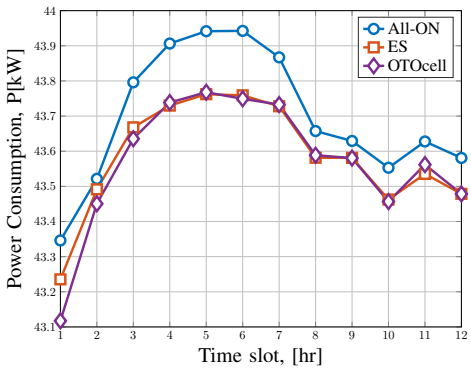

Fig. 2. Power consumption of OTOcell and benchmarks when $N = 4$.

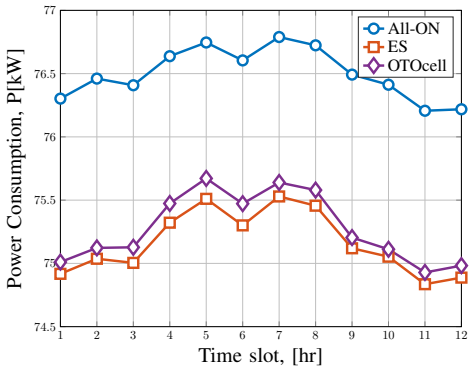

Fig. 3. Power consumption of OTOcell and benchmarks when $N = 12$.

From Table II, when N is 4, it can be observed that the throughput of OTOcell is the same as that of ES, and All-ON approaches, which means than it is able to guarantee the QoS of the network. However, when N increases to 12, a slight decrease in network throughput is observed with OTOcell compared to ES at certain time instances. This can be traced to inappropriate switch ON/OFF decisions due to wrong predictions from the proposed framework occasioned by the increase in network dimension and complexity which makes it more difficult to accurately train the ANN model. Hence, the QoS of the network is slightly reduced when the network dimension increases, but the overall effect on the network is very minimal as revealed in the average throughput (TP avg.) values.

V. Conclusion

In this paper, a UDN with four types of SCs and two deployment scenarios is considered. A cell switching mechanism based on ANN is proposed to determine the optimal cell switching pattern per time instance based on the traffic loads of both the MC and SCs. The simulation results reveals that a significant amount of power savings can be achieved with the proposed model. In addition, the performance of the proposed OTOcell framework is very close to the optimal ES-based solution in terms of power consumption minimization with very minimal effect on the QoS. Future work will consider using more realistic traffic model, carry out complexity, and error probability analysis on the proposed model in order to further ascertain its suitability for UDNs.

TABLE II
Throughput of OTOcell and benchmarks

N	Method	Throughput per time slot [Mbps]						
		4	8	12	16	20	24	TP avg.
4	All-ON	2.55	2.49	2.49	2.48	2.53	2.45	2.48
	ES	2.55	2.49	2.49	2.48	2.53	2.45	2.48
	OTOcell	2.55	2.49	2.49	2.48	2.53	2.45	2.48
12	All-ON	6.45	6.53	6.57	6.68	6.46	6.41	6.55
	ES	6.45	6.53	6.57	6.68	6.46	6.41	6.55
	OTOcell	6.44	6.51	6.55	6.68	6.44	6.40	6.54

VI. Acknowledgement

This work was supported partly by the EPSRC (GCRF) funds (Grant no. EP/P028764/1) and Ajman University (Grant no. 2019-IRG-ENIT-8). The first author was supported by the Nigerian Tertiary Education Trust Fund (TETfund).

References

[1] M. Feng, S. Mao, and T. Jiang, "Base Station ON-OFF Switching in 5G Wireless Networks: Approaches and Challenges," *IEEE Wireless Communications*, vol. 24, no. 4, pp. 46–54, Aug 2017.

[2] A. Shojaeifard, K. Wong, K. A. Hamdi, E. Alsusa, D. K. C. So, and J. Tang, "Stochastic Geometric Analysis of Energy-Efficient Dense Cellular Networks," *IEEE Access*, vol. 5, pp. 455–469, 2017.

[3] J. Wu, S. Jin, L. Jiang, and G. Wang, "Dynamic switching off algorithms for pico base stations in heterogeneous cellular networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, p. 117, 2015.

[4] Ruhong Zeng, Shixiang Zhu, Hongwen Yang, and Jiaxin Zhu, "An artificial neural network based cell switch-off algorithm in cellular system," in *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, Oct 2016, pp. 1434–1439.

[5] A. I. Abubakar, M. Ozturk, S. Hussain, and M. A. Imran, "Q-Learning Assisted Energy-Aware Traffic Offloading and Cell Switching in Heterogeneous Networks," in *2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, Sep. 2019, pp. 1–6.

[6] N. Bui, M. Cesana, S. A. Hosseini, Q. Liao, I. Malanchini, and J. Widmer, "A Survey of Anticipatory Mobile Networking: Context-Based Classification, Prediction Methodologies, and Optimization Techniques," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1790–1821, thirdquarter 2017.

[7] F. Hussain, S. A. Hassan, R. Hussain, and E. Hossain, "Machine Learning for Resource Management in Cellular and IoT Networks: Potentials, Current Solutions, and Open Challenges," *arXiv preprint arXiv:1907.08965*, 2019.

[8] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *arXiv preprint arXiv:1902.02647*, 2019.

[9] L. Wang, S. Chen, and M. Pedram, "Context-driven power management in cache-enabled base stations using a Bayesian neural network," in *2017 Eighth International Green and Sustainable Computing Conference (IGSC)*, Oct 2017, pp. 1–8.

[10] I. Donevski, G. Vallero, and M. A. Marsan, "Neural Networks for Cellular Base Station Switching," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2019, pp. 738–743.

[11] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40–49, October 2011.

[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.