Dynamic Energy-Efficient User Plane Function Selection in 5G Networks

1st Shima Afshar Borji
DITEN - University of Genoa
Genoa, Italy
shima.afshar.borji@edu.unige.it

2nd Roberto Bruschi
DITEN - University of Genoa
CNIT – S2N National Lab
Genoa, Italy
roberto.bruschi@unige.it

3rd Cristina Emilia Costa
CNIT – S2N National Lab
Genoa, Italy
ccosta@cnit.it

4th Chiara Lombardo
DITEN - University of Genoa
CNIT – S2N National Lab
Genoa, Italy
chiara.lombardo@unige.it

***Abstract*—User Plane Function (UPF) is the most demanding component in the 5G Core for computational resources and energy consumption. Static UPF deployments struggle with dynamic traffic patterns, leading to significant energy waste during off-peak hours or potential performance bottlenecks during peak times. This paper proposes a framework for dynamic selection of heterogeneous UPF implementations that leverages traffic prediction and UPF profiling to optimize energy efficiency while maintaining service performance. We profile two distinct UPF implementations characterizing their power consumption and performance across various offered loads. A lightweight Long Short-Term Memory (LSTM) model is employed to predict near-term traffic load based on historical patterns. A utility function, weighting both predicted performance score and power efficiency, guides the selection between the two UPFs. Switching logic incorporating hysteresis and hold-down timers are exploited to minimize excessive transitions. Simulation results over a 7-day period using scaled historical traffic data demonstrate that the proposed dynamic strategy can achieve energy savings of approximately 10.46% compared to the best static UPF deployment without significantly impacting performance. The framework provides a practical approach to enhancing the energy efficiency and sustainability of 5G network operations.**

***Index Terms*—5G, User Plane Function (UPF), Energy Efficiency, Network Function Virtualization (NFV), Network Traffic Forecasting, LSTM, Dynamic Resource Allocation.**

I. Introduction

The implementation of 5G and beyond networks promises significant connectivity with higher data rates, lower latency, and increased capacity. However, this advancement comes at the cost of significantly increased energy consumption within the network infrastructure [1]. The growing number of base stations, the deployment of edge computing resources, and the complexity of the 5G Core (5GC) contribute substantially to this energy footprint, causing challenges for operational costs and environmental sustainability [2]. Enhancing energy efficiency is essential for the sustainable operation of future mobile networks.

Within the 5GC, the User Plane Function (UPF) is a critical component responsible for packet routing, forwarding, Quality of Service (QoS) handling, and usage reporting. Its central role in data plane processing makes it a significant contributor to overall core network energy consumption [3]. The challenge is amplified by the highly dynamic nature of mobile network traffic, which varies significantly based on the time of day, user mobility, and specific events. Many current deployments utilize static UPF configurations, often dimensioned for peak loads. This leads to substantial energy waste during periods of low traffic when resources are underutilized.

Furthermore, the 5G ecosystem allows for diverse UPF implementations, ranging from software-based user-space applications to high-performance kernel-bypass solutions utilizing frameworks like the Data Plane Development Kit (DPDK) and potentially programmable P4-based options [4]. Each implementation presents a unique trade-off between performance and energy consumption. Static allocation fails to leverage these differences dynamically based on real-time network conditions. While some research has explored dynamic UPF placement and chaining [5], [6], a practical framework integrating prediction with selection between distinct implementation types based on energy-performance profiles needs further investigation.

To address these challenges, this paper proposes and evaluates a framework for dynamic, energy-efficient UPF selection. Our approach combines:

ISBN 978-3-903176-72-0 © 2025 IFIP

1) **UPF Profiling:** Experimental assessment of power consumption and performance metrics for different UPF implementations (OpenAirInterface (OAI) user-space and SD-Core's DPDK-based) under varying offered traffic loads.
2) **Traffic Forecasting:** Utilizing a lightweight Long Short-Term Memory (LSTM) model trained on historical data to predict near-term traffic demand.
3) **Utility-Based Decision Engine:** Employing a utility function that weights performance (derived from multiple QoS metrics) and power efficiency (derived from power consumption models) to determine the optimal UPF for the predicted traffic.
4) **Controlled Selection Logic:** Integrating hysteresis margins and hold-down timers to prevent excessive switching while considering the non-negligible costs associated with restarting UPF instances.

We simulate this framework using scaled historical traffic data and evaluate its effectiveness in terms of energy savings, performance impact, and switching frequency compared to static baseline scenarios. The results demonstrate the potential for significant energy reduction through intelligent and adaptive UPF management.

The remainder of this paper is organized as follows: Section II provides the related work. Section III details the proposed framework, including profiling, modeling, prediction, and the decision algorithm. Section IV presents the simulation results and analysis. Finally, Section V concludes the paper and discusses future work.

II. Background and Related Work

The UPF can be implemented in different ways, now usually in the form of lightweight virtualized entities (i.e., containers). Several approaches exist for configuring UPFs, each with specific trade-offs in performance, flexibility, and energy consumption [4]:

- **User-Space Applications:** Often developed for flexibility and ease of deployment, potentially running within standard operating system environments. Solutions like the OAI core network provide examples [7]. These may have lower performance but potentially lower baseline energy consumption compared to highly optimized solutions.
- **DPDK:** These implementations bypass the kernel's networking stack to achieve high packet processing throughput and low latency. Projects like SD-Core provide DPDK-based UPF models [8]. They typically offer superior performance at high loads but may consume more power, especially at idle or low loads, due to polling mode drivers that continuously utilize dedicated CPU cores at 100% workload.
- **Hardware Accelerated Data Planes (P4):** P4 allows for programming the forwarding behavior of network devices, enabling high performance and flexibility. However, P4-based UPFs are not yet common in practical deployments due to their reliance on specialized hardware [4].

These different implementations can be managed in a static or dynamic way depending on whether the network needs to respond to changing traffic conditions. Dynamic approaches allow for optimizing a key performance indicator (KPI), such as energy consumption, throughput, or cost. For this reason, previous studies have recognized the potential benefits of dynamically managing UPF resources. Work by Alevizaki et al. [5] explored dynamic UPF allocation enabled by optical network nodes. Leyva-Pupo and Cervelló-Pastor [9] addressed the placement and chaining problem of UPFs. While these studies highlight the need for flexible UPF management, they mainly focus on network-wide resource orchestration rather than the practical selection between available UPF software implementations with distinct energy-performance characteristics. Our work specifically targets this practical selection challenge, using traffic prediction to choose between implementations.

III. Methodology

This section details the framework developed to dynamically select UPF implementations (OAI User-Space vs. SD-Core DPDK-based) in a 5G network environment, aiming to minimize energy consumption while maintaining acceptable performance levels. The methodology integrates traffic prediction, UPF performance/power profiling, a utility-based decision engine, and simulation-based evaluation and optimization.

A. System Setup

To establish the performance and energy characteristics of different UPF implementations, empirical data was collected within a fully operational 5G test environment deployed on bare-metal servers. This infrastructure featured:

- **SD-Core UPF:** A high-throughput, DPDK-based UPF from the SD-Core project, leveraging hardware acceleration on an AMD EPYC 9474F processor with 512GB RAM and a 100Gbps Network Interface Card (NIC). This corresponds to the "DPDK-UPF" referred to in this study.
- **User-Space UPF:** A software-based UPF deployed by OpenAirInterface Core, handling packets in user-space on an Intel Xeon E5-2680 v4 processor with 256GB RAM and a 40Gbps NIC. This corresponds to the "OAI-UPF" referred to in this study.

Traffic loads were generated using Keysight LoadCore [10], leveraging two dedicated agents to simulate network behavior. The first agent, acting as the Next-Generation Radio Access Network (NGRAN) simulator (as shown in Figure. 1), emulated User Equipment (UE) and Radio Access Network (RAN) behavior, while the second agent functioned as the Data Network (DN) simulator. This setup enabled stress-testing of UPF performance across diverse scenarios, with traffic loads systematically varied. Power consumption was monitored via Scaphandre [11], while network performance metrics (including throughput, uplink/downlink delay, uplink/downlink jitter, and packet loss) were collected from the LoadCore platform to correlate energy usage with operational efficiency.

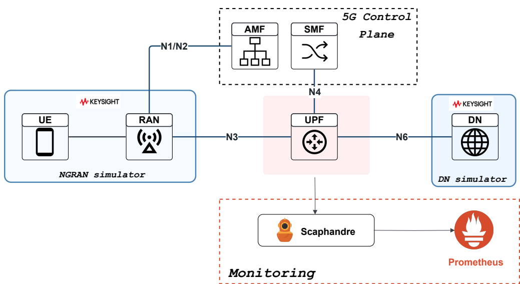

Fig. 1. Simplified Testbed Setup Diagram.

B. Performance Score Formulation

To quantify the overall QoS, a composite `Performance Score`, $P_u(T)$; ranging from 0 to 100, with higher values indicating superior quality, was formulated. This score combines multiple performance metrics, each normalized and weighted according to its importance.

Each metric M was normalized to a 0–100 scale based on its observed minimum (Min_M) and maximum (Max_M) values across the profiling data for both UPFs.

- For "higher-is-better" metrics (e.g., throughput):

$$\text{Norm}_M = 100 \times \frac{M - \text{Min}_M}{\text{Max}_M - \text{Min}_M} \tag{1}$$

- For "lower-is-better" metrics (e.g., delay, loss, jitter, CPU):

$$\text{Norm}_M = 100 \times \frac{\text{Max}_M - M}{\text{Max}_M - \text{Min}_M} \tag{2}$$

Each normalised metric was assigned a weight (w_M) reflecting its contribution to overall performance ($\sum w_M = 1.0$). For each UPF type $u \in \{\text{OAI-UPF}, \text{DPDK-UPF}\}$, we define the final score as:

$$P_u(T) = \sum_{M_u} (w_{M_u} \times \text{Norm}_{M_u}(T)) \tag{3}$$

where the sum is over all selected M_u evaluated at throughput T using the polynomial models. Initially, certain particular weights were used; later, they were optimised.

C. Energy Efficiency Score

Power consumption (W) was normalised across the global minimum ($W_{\text{min,global}}$) and maximum ($W_{\text{max,global}}$) power values observed for both UPFs within the relevant operational throughput range. Lower power yields a higher score:

$$E_u(T) = 100 \times \frac{W_{\text{max,global}} - W(T)}{W_{\text{max,global}} - W_{\text{min,global}}} \tag{4}$$

This score was clipped to $[0, 100]$.

D. Utility Function Definition

To directly model the trade-off between performance and energy efficiency, a `Utility` function was defined. This combines the `Performance Score` with a normalized `Energy Efficiency Score`.

The overall utility is a weighted sum:

$$\mathcal{U}_u(T) = \alpha \cdot P_u(T) + \beta \cdot E_u(T) \tag{5}$$

where $\alpha, \beta \in [0, 1]$: Weights for performance vs. energy ($\alpha + \beta = 1$) and represent prioritisation between performance or power efficiency.

E. Traffic Data and Preprocessing

The historical network traffic data used in this study is sourced from the NetMob 2023 dataset [12], which offers high-resolution, service-level mobile traffic information collected across 20 metropolitan areas in France over a 77-day period in 2019. This dataset provides traffic demand measurements at 15-minute intervals, capturing real-world patterns of various mobile applications. For this analysis, the focus is on video streaming services, such as Netflix.

Given that the dataset contains normalised traffic values—lacking explicit units to preserve confidentiality—the time series data were scaled to align with the minimum capacity of the UPF. This scaling assumes that only traffic manageable by the UPF with the lowest capacity will be steered, ensuring consistency between the dataset's magnitude and the system's operational thresholds.

F. Traffic Forecasting Model

To predict near-term traffic demand, an LSTM model was developed and employed. The model architecture consists of

- An LSTM layer with 50 units, designed to process sequences of historical traffic data. Based on the training data shape, input sequences representing the past 24 hours (96 intervals of 15 minutes) were used.
- A dropout layer following the LSTM layer to mitigate overfitting.
- A final dense (fully connected) layer with a single output unit, providing the traffic prediction for the next 15-minute interval.

It was trained on scaled real traffic data derived from the NetMob 2023 dataset (as described in Section III-E), using past traffic patterns to predict the load for the subsequent interval. Performance during training was monitored using metrics such as Mean Absolute Error (MAE). This predictive model provides the crucial input for our dynamic UPF selection algorithm.

We chose a standard LSTM over more complex variants to maintain low computational overhead and fast inference for near-real-time decision-making. Since the goal of this work is to enable an efficient and responsive UPF selection mechanism, a lightweight LSTM model offered the best trade-off between predictive performance and operational simplicity.

G. Dynamic UPF Selection Algorithm & Simulation

The simulation proceeds interval by interval, corresponding to the 15-minute granularity of the traffic data. In essence, for each interval i the three steps in the following are performed.

Traffic Prediction: The LSTM model forecasts the expected traffic load ($T_{\text{predicted}}$) for the upcoming interval based on the historical data observed up to interval $i-1$. This employs a rolling forecast methodology, where the input sequence for the model is updated with the actual observed value for the interval i before predicting the interval $i+1$.

Candidate UPF Selection: The decision engine evaluates the utility function $\mathcal{U}(T_{\text{predicted}})$ for both OAI-UPF and DPDK-UPF using the predicted traffic and the optimal weights α, β found during the optimization phase (Section IV-C). The UPF implementation resulting in the higher utility score is determined as the theoretically ideal UPF ($\text{UPF}_{\text{ideal}}$) for the predicted conditions.

Selection Control Mechanism: To prevent excessive switching due to prediction noise or traffic fluctuations near the decision threshold and to account for the overhead of UPF transitions, two control mechanisms are applied before initiating a switch from the currently active UPF ($\text{UPF}_{\text{current}}$) to $\text{UPF}_{\text{ideal}}$:

- *Hold-down Timer:* A switch is only considered if the $\text{UPF}_{\text{current}}$ has been active for a minimum duration of I_{hold} intervals. This prevents rapid oscillations.
- *Hysteresis Margin:* A switch is only triggered if the predicted traffic $\hat{T}_i$ not only crosses the utility crossover threshold ($T_{\text{crossover}}$), but also exceeds it by a defined margin (m_{hyst}). This margin helps filter out small fluctuations due to prediction noise.
  The hysteresis margin is defined as:
  $$m_{\text{hyst}} = k \cdot \text{MAE}_{\text{in}} \tag{6}$$
  where k is a tunable sensitivity constant and MAE_{in} is the in-sample Mean Absolute Error of the traffic predictor.

The switching decision rule then becomes:

$$\text{UPF}_{\text{ideal}} = \begin{cases} \text{DPDK-UPF}, & \text{if } \hat{T}_i > T_{\text{crossover}} + m_{\text{hyst}} \\ \text{OAI-UPF}, & \text{if } \hat{T}_i < T_{\text{crossover}} - m_{\text{hyst}} \end{cases} \tag{7}$$

H. Evaluation Metrics

The dynamic strategy is compared against static always-OAI and always-DPDK baselines using total energy consumed (Wh), average performance score, energy saving (%), performance change (%) and switch count. Comparisons are relative to the most energy-efficient static baseline. Additionally, the simulation (Section III-G) was run iteratively across a range of β as thoroughly explained in Section IV-C, to determine the optimal trade-off between performance and energy efficiency.

IV. Results

This section presents the results obtained by applying the methodology described in Section III.

A. UPF Profiling and Model Fitting

The profiling confirmed distinct energy-performance characteristics for the User-space UPF and DPDK-based UPF implementations. Polynomial regression generated functions that accurately modelled the observed data. Key observations include:

- **Performance:** DPDK-UPF generally offered lower average delay, particularly at higher offered loads. OAI-UPF's delay increased more significantly as load increased. OAI-UPF can only handle loads up to 600 Mbps. In terms of CPU utilization , as expected, DPDK-UPF is always at 100% , while OAI-UPF CPU load increases as the traffic load increases.
- **Performance Score:** The overall performance score reflected these trade-offs, with OAI-UPF typically scoring slightly higher at low loads (below approx. 27 Mbps) and DPDK-UPF scoring higher at high loads.
- **Power Consumption:** OAI-UPF exhibited significantly lower idle power and lower power consumption at low offered loads (below approx. 90 Mbps). DPDK-UPF showed higher idle power but scaled more efficiently at higher throughputs, consuming less power.

Figures 2 and 3 visualize these trends. Polynomial power and performance scores are plotted against offer loads for each UPF. The DPDK-UPF is ideal for high-throughput applications due to its stability and low variance. The OAI-UPF is more efficient at idle or light loads but volatile and inefficient at higher loads, confirming their different operational profiles.

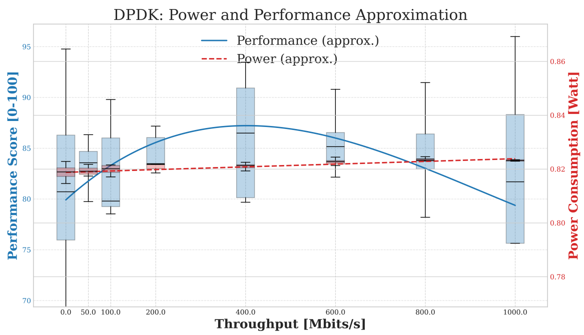

Fig. 2. Empirical measurements and polynomial approximations of Performance (deg. 3) and Power consumption (deg. 1) for DPDK-UPF

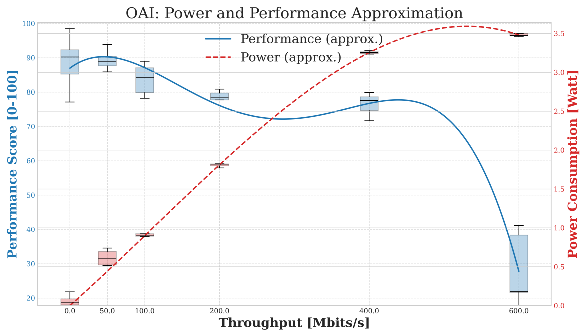

Fig. 3. Empirical measurements and polynomial approximations of Performance (deg. 4) and Power consumption (deg. 4) for OAI-UPF

B. Traffic Forecasting

The LSTM model was trained on 2688 intervals of scaled historical data. Figure 4 illustrates the model's one-step-ahead predictions against the actual scaled traffic during the test period. MAE for these predictions was 25.04 Mbps. The in-sample MAE on the training set, used for setting the hysteresis margin, was 24.70 Mbps.

C. Optimal Utility Weights and Crossover

The optimization process iterated through β values from 0 to 1 with a 0.1 step. Maximizing energy savings while satisfying the performance constraint (Avg. Perf. Score ≥ 70 / Perf. Change $\geq -5\%$), yielded the optimal weights:

- Optimal `Weight_Performance` (α): 0.2
- Optimal `Weight_Power` (β): 0.8

For these weights, the utility crossover point ($T_{\text{crossover}}$) was found to be **95.6** Mbps.

D. Simulation Results with Optimal Weights

The dynamic UPF selection algorithm was simulated over the 672 interval (168 hours) test period using the optimal utility weights ($\alpha = 0.2$, $\beta = 0.8$), the LSTM predictor, a hysteresis margin of 26 Mbps, and a hold-down timer of 4 intervals (1 hour). Table I summarizes the key findings compared to static baselines. Figure 4 shows the traffic, predictions, and dynamic UPF selection over the simulation period.

The results demonstrate that the dynamic strategy achieved an energy saving of **10.46**% compared to the most efficient static baseline (DPDK-UPF). This saving was accompanied by a performance increase of **2.36**%. The system performed 14 switches over the 168-hour (7 days) simulation period (an average of two switches per day). The visualization in Figure 4 shows that the majority of switches occurred during early morning hours, typically between 3:00 AM and 8:00 AM.

E. Sensitivity Analysis

The optimization results show the sensitivity to utility weights. Prioritizing performance (higher α) reduced energy savings but maintained higher average performance scores. Although the performance change for these UPFs is not significant and remains less than 5% even while solely prioritizing power efficiency ($\beta = 1$), even showing a positive change compared to the best baseline, the results are shown for said weights in Figures 5 and 6. Sensitivity to selection control parameters was also observed; for instance, increasing the hold-down timer reduced the switch count but potentially delayed transitions to a more energy-efficient UPF, slightly decreasing overall savings. The chosen optimal parameters represent a specific balance based on the defined performance constraints.

V. Conclusion and Future Work

This paper proposed a dynamic, energy-efficient UPF selection framework for 5G networks using traffic prediction and UPF profiling. We compared user-space OAI-UPF and DPDK-UPF, modeled their energy-performance trade-offs with polynomial functions, and used a lightweight LSTM to forecast traffic. A utility function balancing performance and energy guided the switching, supported by practical controls such as hysteresis and hold-down period.

Simulation results show that dynamic UPF selection can save **10.46%** energy with a slight performance improvement of **2.36%** compared to static deployments.

As future work, we plan to consider the real-world impact of switching between UPFs, such as momentary packet loss and the extra power used when starting or stopping containers, so that these effects are reflected in the selection process and lead to even more practical decisions.

Overall, this work presents a practical methodology and highlights the potential of intelligent UPF selection to enhance energy efficiency in 5G core networks.

Acknowledgment

This work has been partially supported by the Horizon Europe Research and Innovation Action **6Green** (Grant

TABLE I
SIMULATION RESULTS SUMMARY WITH OPTIMAL UTILITY WEIGHTS (α=0.2,β=0.8)

Metric	***Dynamic Strategy***	***Always-OAI Baseline***	***Always-DPDK Baseline***	***Comparison vs Best*** (DPDK-UPF)
Total Energy (Wh)	123.29	225.63	137.69	**10.46% Saving**
Avg. Performance Score	86.15	82.04	84.17	**+2.36% Change**

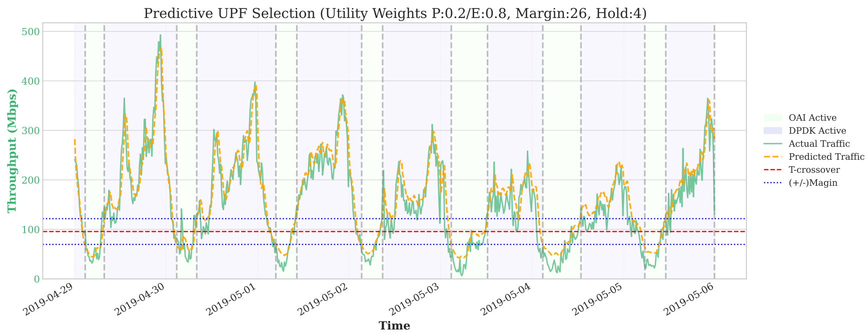

Fig. 4. Actual traffic (from the NetMob dataset) and predicted traffic over the 7-day test period. The selected UPF (in the background), the T-crossover, and the upper and lower margins are reported aswell.

Agreement no. 101096925), by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, within the partnership on **"Telecommunications of the Future"** (PE00000001 – program **"RESTART"**), and the project **"SoBigData.it – Strengthening the Italian RI for Social Mining and Big Data Analytics"** (Prot. IR0000013 – Call n. 3264 of 12/28/2021).

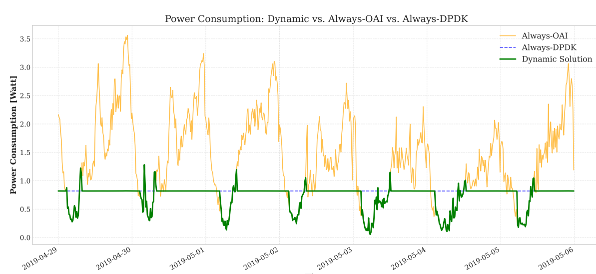

Fig. 5. Visualization of dynamic solution's power consumption compared to static baseline.

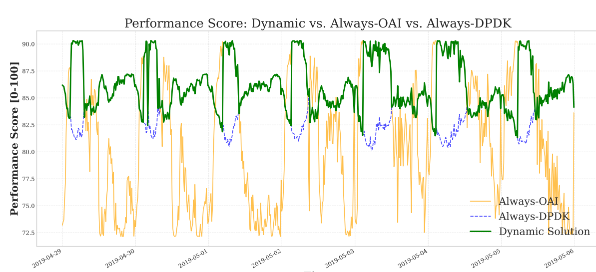

Fig. 6. Visualization of dynamic solution's performance score compared to static baseline.

REFERENCES

[1] Y. Ding, H. Duan, M. Xie, R. Mao, J. J. Wang, and W. Zhang, "Carbon emissions and mitigation potentials of 5g base station in china," *Resources, Conservation and Recycling*, vol. 182, p. 106330, 2022.

[2] A. Mughees, M. Tahir, M. A. Sheikh, and A. Ahad, "Towards energy efficient 5g networks using machine learning: Taxonomy, research challenges, and future research directions," *IEEE Access*, vol. 8, pp. 187 498–187 522, 2020.

[3] A. Bellin, N. D. Cicco, D. Munaretto, and F. Granelli, "Power consumption-aware 5g edge upf selection using deep reinforcement learning," in *Proceedings of the IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, 2024, pp. 1–6.

[4] J. Rischke, C. Vielhaus, P. Sossalla, J. Wang, and F. H. P. Fitzek, "Comparison of upf acceleration technologies and their tail-latency for urllc," in *2022 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*. IEEE, 2022, pp. 146–151.

[5] V. M. Alevizaki, A. I. Manolopoulos, M. Anastasopoulos, and A. Tzanakaki, "Dynamic user plane function allocation in 5g networks enabled by optical network nodes," in *2021 European Conference on Optical Communication (ECOC)*. IEEE, 2021, pp. 1–4.

[6] I. Leyva-Pupo and C. Cervelló-Pastor, "Efficient solutions to the placement and chaining problem of user plane functions in 5g networks," *Journal of Network and Computer Applications*, vol. 197, p. 103253, 2022.

[7] OpenAirInterface, "Openairinterface: 5g software alliance for democratising wireless innovation," https://openairinterface.org.

[8] O. N. F. (ONF), "Sd-core," https://opennetworking.org/sd-core/.

[9] I. Leyva-Pupo and C. Cervelló-Pastor, "An intelligent scheduling for 5g user plane function placement and chaining reconfiguration," *Computer Networks*, 2023.

[10] Keysight, "P8900s loadcore – core network solutions," https://www.keysight.com/us/en/product/P8900S/loadcore-core-network-solutions.html.

[11] B. Petit, "scaphandre," 2023. [Online]. Available: https://github.com/hubblo-org/scaphandre

[12] O. E. Martínez-Durive, S. Mishra, C. Ziemlicki, S. Rubrichi, Z. Smoreda, and M. Fiore, "The netmob23 dataset: A high-resolution multi-region service-level mobile data traffic cartography," 2023.