Energy Consumption of NFV-MEC-Based 6G Networks Supporting Different Services

Caio Souza, Marcos Falcão, Peterson Yoshioka, Andson Balieiro, Elton Alves

***Abstract*— The Sixth Generation of Mobile Networks (6G) supports Immersive Communication (IC) and Hyper Reliable and Low-latency Communication (HRLLC). To accommodate heterogeneous services with contrasting requirements, Multi-Access Edge Computing (MEC), Network Function Virtualization (NFV), and Network Slicing (NS) are essential in 6G networks. Their combination provides multiple customized virtual networks and offers flexibility and quick network resource allocation, allowing virtualized network functions (VNFs) to run on generic hardware near the user and resources to be scaled according to demand dynamics. Besides supporting different services, achieving sustainability and energy efficiency are key considerations for 6G networks. Thus, this paper presents a Continuous-Time Markov Chain (CTMC)-based model to analyze the energy consumption of NFV-MEC nodes hosting two service types (HRLLC and IC), considering the overhead introduced by virtualization. Additionally, to cope with sudden load variations and address the strict latency requirements of HRLLC services, Dynamic Resource Allocation (DRA) and service prioritization are also incorporated into the NFV-MEC node. The energy consumption is analyzed under different NFV-MEC nodes and service characteristics such as resource amounts, failure rates, setup times, and service rates. The model may assist network operators in properly dimensioning/configuring the MEC-NFV node according to the energy budget.**

***Keywords*— NFV, MEC, 6G Networks, Energy Consumption, CTMC**

I. Introduction

The Sixth Generation of Mobile Networks (6G) will become a reality around 2030 [1] and promises higher data rate, reliability, near-zero latency, and connectivity beyond the Fifth Generation (5G) capabilities, supporting advanced applications such as holographic communication, advanced Augmented Reality (AR)/Virtual Reality (VR) experiences, tactile internet, digital twins, and AI-driven autonomous systems [2]. 6G is expected to expand the 5G services, including Immersive Communication (IC) and Hyper Reliable and Low-latency Communication (HRLLC). The former boosts the enhanced Mobile Broadband (eMBB) scenario and provides rich and interactive mobile services to users, including interactions with machine interfaces. The latter extends the Ultra Reliable and Low Latency Communication (URLLC) and encompasses applications with more stringent requirements on reliability and latency[1].

To accommodate heterogeneous services with contrasting requirements, Multi-Access Edge Computing (MEC), Network Function Virtualization (NFV), and Network Slicing (NS) are essential in 6G networks [3]. MEC provides computing resources closer to the user equipment (UE). Hosting services and applications in MEC nodes can break down user latency to sub-millisecond levels, which is crucial for HRLLC (e.g., autonomous vehicles, remote surgery, and tactile internet). Additionally, MEC helps reduce data traffic towards core and external networks. By decoupling network functions from proprietary hardware, NFV offers flexibility and quick network resource allocation, allowing virtualized network functions (VNFs) to run on generic hardware and resources to be scaled according to demand dynamics. Combined with MEC, the advantages of NFV can also be brought near UEs, benefiting both HRLLC and IC services. NS, in turn, enables the coexistence of different services (e.g., HRLLC and IC) sharing the same physical infrastructure by creating multiple customized virtual networks.

Besides supporting different services, achieving sustainability and energy efficiency are key considerations for 6G networks [4] as these directly impact both the economic and ecological aspects of cellular networks. Energy costs are a significant component of the overall operational expenditure for network operators. Hence, companies need to adopt energy-efficient solutions, not only for reducing their power consumption cost but also taking on social responsibility to reduce the carbon footprint [5].

Although MEC, NFV, and NS are pivotal in addressing this challenge, ensuring that 6G networks can deliver high performance while maintaining energy efficiency, the energy consumption associated with NFV-MEC nodes, considering different types of services hosted on the MEC-NFV node and the overhead introduced by virtualization such as VNF instantiation time and VNF recovery time, must be carefully analyzed, mainly when studies [6] [7] indicate that the distributed and pervasive nature of MEC nodes causes a noticeable usage, increasing their costs, carbon footprint, and energy requirements. By understanding the energy consumption pattern including the virtualization costs, operators can develop more efficient and sustainable NFV-MEC systems.

In this respect, this paper analyzes the energy consumption of NFV-MEC nodes hosting two service types (HRLLC and IC), considering the overhead introduced by virtualization. Additionally, to cope with sudden load variations and address the strict latency requirements of HRLLC services, Dynamic Resource Allocation (DRA) and service prioritization are also incorporated into the analyzed system. These mechanisms scale resources on demand and prioritize HRLLC services during resource allocation, respectively. By designing a Continuous Time Markov Chain-based model, we analyze energy

Caio Souza, Marcos Falcão, Perterson Yoshioka, Andson Balieiro, Centro de Informática (CIn), Universidade Federal de Pernambuco (UFPE), Brasil, e-mail: {cbbs,mrmf,pmsy, amb4}@cin.ufpe.br; Elton Alves, Universidade do Sul e Sudeste do Pará, Marabá, Brasil, e-mail: eltonalves@unifesspa.edu.br.
This study was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Code 001

consumption under different node and service characteristics such as resource amounts, failure rates, setup times, and service rates. The model may assist network operators in properly dimensioning/configuring the MEC-NFV node according to the energy budget.

The remainder of this paper is organized as follows. Section II discusses some works. The system model and power consumption formulation are presented in III. Results and Analysis are conducted in IV. Section V concludes this paper.

II. Related Work

Sustainability and energy efficiency stand as pillars of 6G networks, as they directly impact both the economic and ecological aspects of cellular networks and are linked to the United Nations' Sustainable Development Goals (UN SDGs) for 2030 [4]. Thus, efforts have to be devoted to minimizing environmental impact by optimizing network architectures, reducing power consumption, and developing energy-efficient solutions in different network segments. In this respect, the authors in [8] propose a strategy for balancing task delay requirements and energy consumption in Integrated Sensing and Communications (ISAC)-aided 6G V2X networks using MEC. They aim to minimize queuing latency with long-term latency and energy consumption constraints for data fusion computing tasks and adopt the Lyapunov optimization method. Although their joint computation offloading and resource allocation (JCORA) scheme presents great results, it neglects failure events during task processing and the different service categories coexisting in MEC nodes, which are expected in real 6G networks.

In [9], a Resource-Ability Assisted Service Function Chain (SFC) Embedding and Scheduling algorithm for virtualization-based 6G Networks is proposed. To enhance the SFC embedding and scheduling, the solution selects nodes with strong capabilities and sufficient resources to accommodate SFCs. Results show that the algorithm achieves a higher SFC acceptance ratio compared to previous methods. [10], in turn, addresses the Virtual Network Function (VNF) placement in Non-Terrestrial Networks (NTNs), considering resource constraints and bandwidth limitations. It focuses on delay-aware VNF placement to support ultra-low delay services and improve resource utilization. The authors propose Linear Programming-based and Hungarian-based solutions to deal with VNF placement problem, achieving superior performance in terms of resource utilization and execution time. Although both works present important findings, they overlook significant factors in its analysis and formulation, including VNF failure, setup time, and energy consumption. These aspects are essential for designing and operating 6G networks that meet reliability, efficiency, and sustainability expectations.

Regarding the dynamic allocation in NFV-MEC nodes, our previous works [11] [12] propose CTMC-based models to analyze MEC-NFV node performance when supporting URLLC services, considering VNF failure and setup/repair time. Additionally, the former encompasses a resource pre-initialization strategy to mitigate the negative effects of VNF failures and setup time [11] in a container-based environment. While the latter [12] evaluates URLLC services running in a hybrid NFV-MEC node that leverages the strengths of both virtual machines and container technologies. It allows a service provider to properly dimension a MEC-enabled UAV node under availability, power consumption, reliability, and latency perspectives. But, the current paper differs from both, as it focuses on the energy consumption analysis of the NFV-MEC node supporting two 6G service categories, IC and HRLLC, with a prioritization strategy being employed.

III. NFV-MEC System

A. System Overview

In this work, we focus on a single isolated NFV-MEC node designed to support two types of services: IC (green flow) and HRLLC (orange flow), as shown in Fig. 1. Service requests originate from UEs and are forwarded by the Radio Access Network (RAN) to the NFV-MEC node, which dynamically scales containerized VNFs on demand to process them. The containers are software unities comprising source code, libraries, dependencies and offer portable, isolated environments for running applications, making them suitable for supporting VNFs scaling for critical services [13]. The system setup includes a finite number of containers and limited buffer capacity for each service type. As a result, incoming requests might need to wait in a buffer until a container becomes available.

Each container is responsible for executing a single VNF independently, and a central unit manages the admission control of new requests based on resource availability. When resources are available (containers or buffer positions), a request is admitted and placed in the buffer if all containers are occupied or assigned to a container otherwise. This mechanism ensures efficient resource utilization and service request processing within the NFV-MEC node.

Our system encompasses a dynamic VNF auto-scaling strategy to manage load variations during NFV-MEC node operation. Before a containerized VNF is ready for service processing, it must undergo an initialization process, which incurs a setup delay. Additionally, failures may occur during service processing, necessitating repair time. If a containerized VNF fails, it will be restarted, and the request that was being served by it will be reallocated to another VNF if available; otherwise, it will be placed back in its service queue with higher priority over new requests. In either case, the service processing is restarted. This approach ensures resilience and continuity in service processing, maintaining efficient operation despite potential failures.

Given that HRLLC services present strict latency requirements, the resource allocation incorporates a prioritization policy that favors this service type as follows: (1) HRLLC services are prioritized over IC services; therefore, containers being released or activated are first allocated to HRLLC services. (2) If an HRLLC service is waiting to be processed and an IC service is completed, the released container will be restarted with the VNF for HRLLC services. However, if other containers are available, the current container is allocated to the next IC service or deactivated if the IC buffer is empty.

(3) Preemption of the lowest priority IC service in processing is not allowed.

B. System Modeling

Analytical models are a valuable alternative to efficiently evaluate large-scale distributed MEC infrastructure projects, especially when simulations and testbeds are costly or unfeasible. The NFV-MEC system is modeled using an M/N/c/k+K queue with First-Come, First Served (FCFS) service discipline, two user types, prioritization, limited buffers for each user type, failure, and resource setup/repair time. The model states are denoted by the tuple (i, j, l, m), where $i, j, l, m \in N$. Here, i and j represent the number of HRLLC and IC services in the system, respectively, and l and m mean the number of active containers for each user type, with $l + m$ being less than or equal to the maximum number of containers (c). Service request arrivals adopt a Poisson process with rates λ_H for HRLLC services and λ_I for IC services. The c available containers provide service processing with exponentially distributed service times, with rates μ_H for HRLLC and μ_I for IC. Similarly, the container initialization and failure occurrence times adopt exponential distributions with rates α and γ, respectively. The maximum system capacity for HRLLC and IC users are denoted by k and K, respectively. Since the model is defined, the state probabilities of the system in steady-state ($\pi_{i,j,l,m}$) may be computed as in [12] and the power consumption metric derived as described in Section III-C.

C. Power Consumption (PC)

The computational power consumption is an important part of the operational costs and must be considered by the service provider for resource planning the NFV-MEC system. In our formulation, the mean power consumption ($\overline{PC}$) is given by the combination of the mean number of virtual resources and energy consumption constants for each operating state: Setup and Busy. The power consumption (in Watts) of a single container in setup state is denoted as P^{CT}_{setup} while in the busy state is P^{CT}_{busy}. It is worth noting that PC encompasses the power consumption for dealing with both service types.

Eqs. (1) and (2) denote the mean number of containers $\overline{CT}$ in the Busy and Setup states, respectively. The former equation captures the mean amount of containers in the busy state by iterating over each system state with service load and varying the combination of the number of each container type from 0 until the number of services from a particular category or the maximum resources available in the system. Moreover, Eq. (2) calculates the mean number of containerized VNFs in setup by iterating over states where the number of online services is greater than the total number of active resources for each service category. Finally, the total mean power consumption ($\overline{PC}$) is given by Eq. (3).

$$\overline{CT}_{busy} = \sum_{i=0}^{k}\sum_{j=0}^{K}\sum_{l=0}^{min(c,i)}\sum_{m=0}^{min(c-l,j)} (l+m)\pi_{i,j,l,m} \quad (1)$$

$$\overline{CT}_{setup} = \sum_{i=0}^{k}\sum_{j=0}^{K}\sum_{l=0}^{min(c,i)}\sum_{m=0}^{min(c-l,j)} min((c-l-m), (i+j-l-m))\pi_{i,j,l,m} \quad (2)$$

$$\overline{PC} = P^{CT}_{setup}\overline{CT}_{setup} + P^{CT}_{busy}\ \overline{CT}_{busy} \quad (3)$$

IV. Results

The analytical results were validated against extensive discrete-event simulations (Figs. 2-4), where the lines denote the analytical and the markers represent simulation results. We evaluate the energy consumption of the NFV-MEC node considering three scenarios. The first scenario (Section IV-A) analyzes the impact of the coexistence of different user types by adopting multiple IC loads (λ_I). The two subsequent scenarios simultaneously assess the influence of: (Section IV-B) container setup rates (α) and failure rates (γ), which aims to demonstrate the impact of system improvements (hardware and/or software) to reduce the time in which network functions need to get ready to process services, and the influence of using components with different levels of reliability to provide services; HRLLC service rate (μ_H) and IC service rate (μ_I) (Section IV-C), with the objective to show how enhancements in service processing speed, achieved through the utilization of advanced processing units and optimized algorithms, can impact the energy consumption.

In all scenarios, the HRLLC service arrivals (λ_H) ranged from 2.5 to 25 requests/ms in order to analyze the NFV-MEC node under different HRLLC loads. Unless stated otherwise, the baseline values for failure (γ) and setup rates (α) were set to 0.001 and 1 unit/ms, respectively, following [15]. For the power consumption of each container in different operation states, we adopted the values from the network-intensive experiment in [13], which are summarized in Table I. The remaining parameters can be found in Table II, following values defined in 3GPP Release 16 (TR 38.824) [14].

Next sections (IV-A - IV-C) display average results. For every obtained using the analytical model, 10 simulation instances, comprising 27000000 simulation steps and 2200000 services attended each, were conducted. The Bootstrap method [16] was employed, with both resample size and the number of (re)samplings set at 30 and 1000, respectively. This was done considering a 95% confidence level. Bars were omitted due to the negligible difference between upper and lower bounds and to prevent overcrowding of the graphs.

TABLE I: Power Consumption for Container Different States

Parameter	Value
Setup Container (P^{CT}_{setup})	8 W
Busy Container (P^{CT}_{busy})	23 W

A. Effects of Varying the IC Load (λ_I)

In this scenario, the Energy Consumption (Fig. 2) exhibited two different behaviors from $\lambda_H = 2.5$ to $\lambda_H = 10$: an increasing trend for part of the configurations ($\lambda_I = 5$ and $\lambda_I = 10$) and a decreasing trend for the remaining curves.

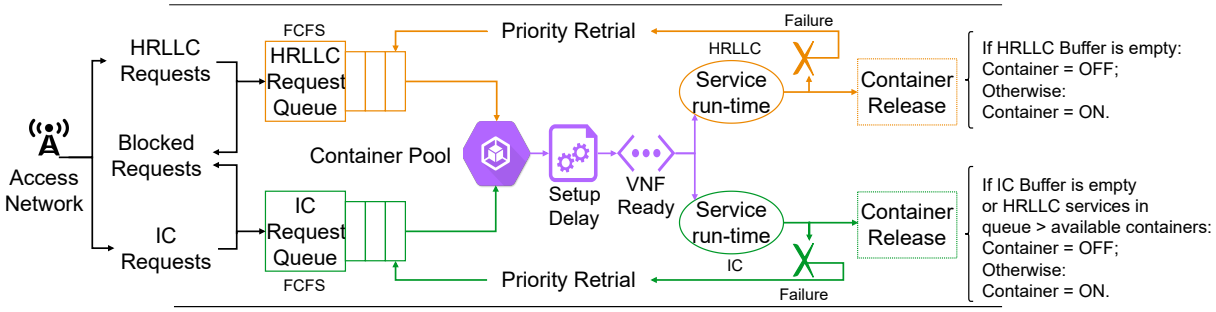

Fig. 1: Resource Allocation Flow in a NFV-MEC node supporting HRLLC and IC services

TABLE II: Evaluation Scenarios

Section	λ_I	α	γ	μ_H	μ_I
IV-A	5,10,15,20,25,30	1	10^{-3}	2	2
IV-B	10	1,2,4	10^{-2}, 10^{-3}	2	2
IV-C	10	1	10^{-3}	1,2,4	1,2

This is due to the summation of the arrival rates of both user types, i.e., when the sum of the arrival rates is lower than the total processing capacity of the system's containers, the curves tend to increase since the idle containers are being activated to meet newly arrived requests. This increase in energy consumption is not directly proportional as we can see in the blue curves, where the increase in λ_I from 5 to 10 at the point where λ_H = 5 (50% considering the sum of both rates) results in an energy consumption increase of 30.6%, while at the point where λ_H = 10 (40% increase) the impact on energy consumption is only 10.2%. Conversely, when these rates exceed the processing capacity of the system, a slight decreasing trend can be observed in the curves. This is attributed to the re-initialization of containers to prioritize HRLLC requests. During container re-initialization, the containers spend more time in setup mode, which uses less energy compared to a processing state, thus resulting in lower energy consumption. The curves tend to converge as the arrival rate of HRLLC requests increases, causing fewer IC requests to be served and subsequently reducing the number of container re-initializations for different service types. As the containers are no longer being reinitialized, they spend more time in the processing state, leading to a new increase in overall energy consumption. But although some curves present a higher energy consumption at points from λ_H = 12.5, this relation presents a better use of energy resources, given that this increase is mostly the result of an increase in the general service rate of the system.

B. Effects of the container setup rate (α) and service failure rate (γ)

Regarding energy consumption of this experiment (Figure 3), higher container setup rates, such as the green/yellow ($\alpha = 2$) and red/orange ($\alpha = 4$) curves, lead to greater energy consumption. This can be attributed to the fact that with higher setup rates, less time is spent in the setup phase, making containers more frequently available. Since the processing phase requires more power compared to the setup phase, the total energy consumption monotonically increases, converging

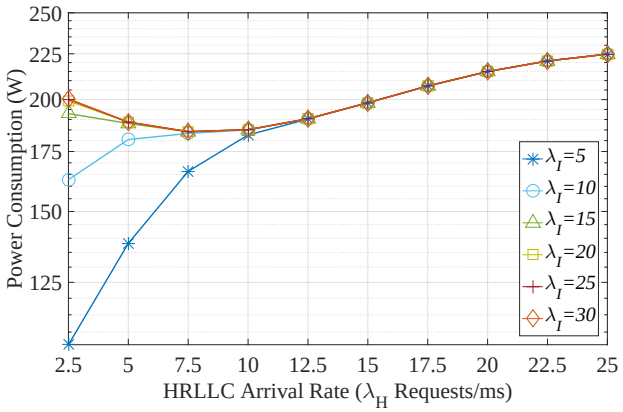

Fig. 2: Power Consumption under Different IC Loads

around $\lambda_H = 25$ to 225. As in the previous experiment, we can visualize the impact of competition for processing resources between service categories when request arrival rates reach $\lambda_H = 7.5$, resulting in a slowdown in the growth of energy consumption for curves with lower setup rates. Additionally, although the impact was small, it is worth noting that curves depicting higher service failure rates exhibit lower energy consumption when comparing the pair of curves with the same α (e.g., light and dark blue lines). However saving energy resources is a good indication, this phenomenon directly impacts the quality of service provided to users, who will have to restart processing their requests. This is due to the increased number of container resets for failed requests, leading to a higher proportion of containers in the setup state.

C. Effects of the HRLLC service rate (μ_H) and IC service rate (μ_I)

Finally, the impacts of varying the service rate in energy consumption are presented in (Fig. 4), once again, a higher service rate for IC users leads to lower energy consumption, especially in the leftmost region of the figure, corresponding to low HRLLC loads, i.e., when the system is predominantly occupied by IC requests. This occurs because a higher IC service rate leads to greater availability for this service category, particularly in the leftmost region of the graph. For configurations with the same values of μ_H, the curve with $\mu_I = 2$ results in greater availability compared to those with $\mu_I = 1$. However, this effect decreases as the HRLLC arrival rate increases, resulting in convergence in the rightmost part of

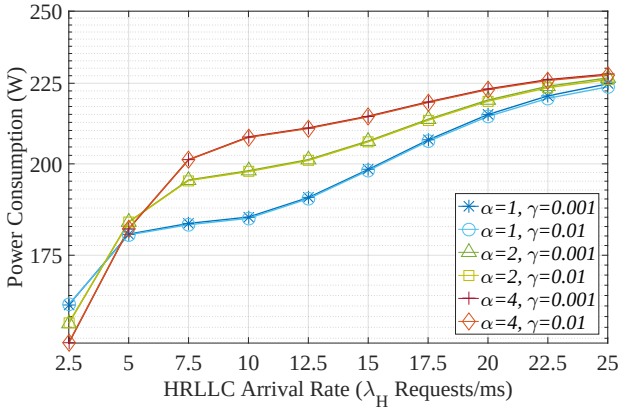

Fig. 3: Power Consumption for Different Setup and Failure rates

the graph. Furthermore, a higher HRLLC service rate implies less time spent by these requests hogging resources, leading to greater availability of processing resources for IC requests. In other words, greater availability caused by higher service rates corresponds to lower energy consumption.

Furthermore, when considering the three different configurations with $\mu_H = 1$, $\mu_H = 2$, and $\mu_H = 4$, significant differences of up to 40 W were observed. For instance, at $\lambda_H = 10$, the configuration with $\mu_H = 4$ and $\mu_I = 2$ (orange line) exhibits a consumption of approximately 175 W, while the configuration with $\mu_H = 2$ and $\mu_I = 2$ (yellow line) consumes around 215 W. This finding is particularly relevant as the experiment maintained the same amount of resources (containers) for all curves, varying only the service rates. In subsequent experiments, different resource and buffer amounts will be analyzed.

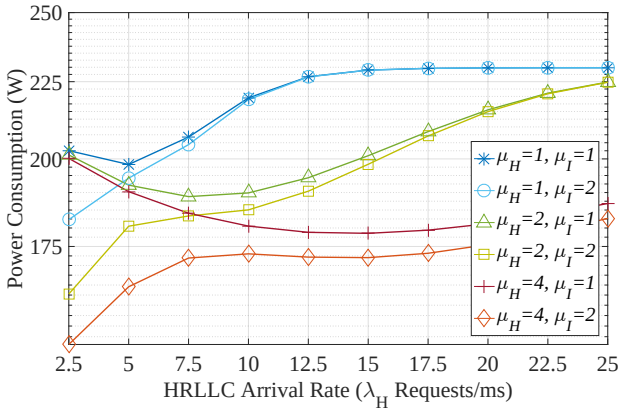

Fig. 4: Power Consumption for Different HRLLC and IC Service Rates

V. Conclusion

This paper analyzed the energy consumption of NFV-MEC nodes hosting two service types (HRLLC and IC), considering the overhead introduced by virtualization, Dynamic Resource Allocation (DRA), and service prioritization. These features were incorporated into a Continuous-Time Markov Chain (CTMC)-based model, and three scenarios with different NFV-MEC node configurations were evaluated. Results showed that higher container setup rates increase energy consumption by reducing the setup time and making containers to more frequently enter the energy-intensive processing phase. Conversely, improved resource utilization and greater availability due to higher service rates can lead to lower overall energy consumption. The proposed model may assist network operators in properly dimensioning/configuring MEC-NFV nodes according to the energy budget. Future directions include designing energy and performance-efficient solutions for NFV-MEC node dimensioning considering different service types.

References

[1] R. Liu, H. Lin, H. Lee, F. Chaves, H. Lim and J. Sköld, "Beginning of the Journey Toward 6G: Vision and Framework," in IEEE Communications Magazine, vol. 61, no. 10, pp. 8-9, October 2023, doi: 10.1109/MCOM.2023.10298069

[2] G. Gui, et al. 6G: Opening new horizons for integration of comfort, security, and intelligence IEEE Wirel Commun, 27 (5) (2020), pp. 126-132.

[3] L. Zhao, G. Zhou, G. Zheng, C. -L. I, X. You and L. Hanzo, "Open-Source Multi-Access Edge Computing for 6G: Opportunities and Challenges," in IEEE Access, vol. 9, pp. 158426-158439, 2021, doi: 10.1109/ACCESS.2021.3130418

[4] Imoize, A.L.; Adedeji, O.; Tandiya, N.; Shetty, S. 6G enabled smart infrastructure for sustainable society: Opportunities, challenges, and research roadmap. Sensors 2021, 21, 1709.

[5] A. Abrol and R. K. Jha, "Power Optimization in 5G Networks: A Step towards GrEEn Communication", IEEE Access, vol. 4, pp. 1355-74, 2016.

[6] B. Joshi, Breaking the energy curve: Why service providers should care about 5G energy efficiency, Feb. 2019, Available: https://www.ericsson.com/en/blog/2019/2/breaking-the-energy-curve-5g-energy-efficiency.

[7] Vertiv and 451 research survey on 5G energy consumption at MWC19, [online] Available: https://www.vertiv.com/en-emea/about/news-and-insights/news-releases/2019/mwc19-vertiv-and-451-research-survey-reveals-more-than-90-percent-of-operators-fear-increasing-energy-costs-for-5g-and-edge/

[8] Q. Liu, R. Luo, H. Liang and Q. Liu, "Energy-Efficient Joint Computation Offloading and Resource Allocation Strategy for ISAC-Aided 6G V2X Networks," in IEEE Transactions on Green Communications and Networking, vol. 7, no. 1, pp. 413-423, March 2023, doi: 10.1109/TGCN.2023.3234263.

[9] H. Cao et al., "Resource-Ability Assisted Service Function Chain Embedding and Scheduling for 6G Networks With Virtualization," in IEEE Transactions on Vehicular Technology, vol. 70, no. 4, pp. 3846-3859, April 2021, doi: 10.1109/TVT.2021.3065967.

[10] Y. Yue et al., "Delay-aware and Resource-efficient VNF placement in 6G Non-Terrestrial Networks," 2023 IEEE Wireless Communications and Networking Conference (WCNC), Glasgow, United Kingdom, 2023, pp. 1-6, doi: 10.1109/WCNC55385.2023.10118893.

[11] C. Souza, M. Falcao, A. Balieiro and K. Dias, "Modelling and Analysis of 5G Networks Based on MEC-NFV for URLLC Services," in IEEE Latin America Transactions, vol. 19, no. 10, pp. 1745-1753, Oct. 2021, doi: 10.1109/TLA.2021.9477275

[12] M. Falcão, C. Souza, A. Balieiro and K. Dias, "Dynamic Resource Allocation for URLLC in UAV-Enabled Multi-access Edge Computing," 2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Gothenburg, Sweden, 2023, pp. 293-298, doi: 10.1109/EuCNC/6GSummit58263.2023.10188346.

[13] R. Morabito, "Power Consumption of Virtualization Technologies: An Empirical Investigation," 2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC), Limassol, Cyprus, 2015, pp. 522-527, doi: 10.1109/UCC.2015.93.

[14] ETSI and 3GPP, "'5G;NR;Physical layer procedures for data (3GPP TS 38.214 version 16.2.0 Release 16",ETSI and 3GPP 2020.

[15] K. Kaur, T. Dhand, N. Kumar and S. Zeadally, "Container-as-a-Service at the Edge: Trade-off between Energy Efficiency and Service Availability at Fog Nano Data Centers," in IEEE Wireless Communications, vol. 24, no. 3, pp. 48-56, June 2017, doi: 10.1109/MWC.2017.1600427

[16] K.Singh and M. Xie, Bootstrap: A Statistical Method, 2008, available at http://www.stat.rutgers.edu/home/mxie/rcpapers/ bootstrap.pdf.