A Preliminary Study on the Power Consumption of Virtualized Edge 5G Core Networks

Arturo Bellin
DISI
University of Trento
Trento, Italy
R&I Department
Athonet S.r.l.
Bolzano Vicentino, Italy
ORCID ID: 0000-0002-4042-847X

Marco Centenaro
R&I Department
Athonet S.r.l.
Bolzano Vicentino, Italy
ORCID ID: 0000-0003-1664-8015

Fabrizio Granelli
DISI
University of Trento
Trento, Italy
ORCID ID: 0000-0002-2439-277X

***Abstract*—Other than pure performance and cybersecurity, a value that is becoming increasingly important for a mobile network is its *power consumption*. In fact, the transition from legacy network deployments tightly coupled with the underlying hardware towards fully virtualized ones yields distinct options based on the adopted virtualization technology, each of which deserve appropriate evaluation in terms of energy efficiency. In this paper, we aim at providing a preliminary assessment of the realistic power consumption of a fifth-generation core network deployed in a network edge environment leveraging bare metal, containers, and virtual machines. The results are based on a testbed consisting of commercial off-the-shelf hardware and open-source software, and show that the deployment based on virtual machines is the first one that saturates the power consumption, thus reducing the maximum achievable throughput. These preliminary insights show the feasibility of a real-time power monitoring system that can condition the dynamic policies applied by the 5G network orchestrator.**

***Index Terms*—5G; core network; energy efficiency; virtualization; network edge; testbed.**

I. INTRODUCTION

Recent studies concerning the environmental sustainability of the information and communication technology sector estimate that, in 2020, it contributed for a share between 1.8% and 2.8% to global greenhouse gas emissions [1].

The widespread adoption of 5G is expected to increase this share even more, owing also to the drastic increase in data traffic, connected devices and use of computing resources in the network. As edge computing and private mobile networks become more popular, it is inevitable that an increasing amount of users and data traffic will be handled by the network edge. While traditionally the attention of manufacturers was mainly paid to the Radio Access Network (RAN) [2], the innovative design of the 5G core network (5GC) is likely to increase its weight in the power consumption balance. In fact, the 5GC Service-Based Architecture (SBA) fully embraced the paradigm of Network Function Virtualization (NFV), allowing for the deployment of network functions on virtualized commercial off-the-shelf (COTS) hardware or on public cloud computing infrastructures instead of application-specific integrated circuit. This made the distributed deployment of virtual network functions (VNFs) at the network edge much easier, enabling new use cases and providing better privacy, latency and reliability. 5GCs entirely or partially deployed on-premises are more likely to leverage smaller and less efficient COTS hardware, without the scaling and consolidation benefits provided by larger data centers. According to a Natural Resources Defense Council report [3], large hyperscalers server infrastructure represented less than 5% of the United States' data center energy use, with the remaining 95% consumed by less efficient small and medium data centers.

The ITU-T has recently defined a key performance indicator (KPI) for the carbon emission intensity of a network focused on the energy consumption with respect to served data traffic, not only encouraging the reduction of network electricity consumption, but also advocating the use of low-carbon energy supply and the improvement of energy utilization efficiency [4].

In this paper, we want to focus on the energy efficiency of 5G systems (5GSs) deployed at the edge, and specifically on their heart, that is, the 5GC.

It has become increasingly difficult to model the power consumption of the 5GC given the heterogeneity of the underlying hardware and virtualization technologies available at the edge. The choice of both hardware and hypervisor strongly influences the overall energy demand. It is not possible to rely only on fixed models and technical specifications from the hardware manufacturers since different networks, or different components within the same network, may be deployed on diverse kinds of infrastructures. It is therefore necessary to make use of dynamic and adaptable mechanisms for the real-time monitoring of the network power consumption, which can be fed to the network orchestrator to enable it take the appropriate operations, administration and management

---

The doctoral studies and research work of Arturo Bellin are jointly supported by Athonet S.r.l. and the Italian National Inter-University Consortium for Telecommunications (CNIT).

The work leading to this publication was partially funded by the European Union's Horizon Europe under Grant Agreement no. 101096342 (HORSE project) and Grant Agreement no. 101096925 (6Green project).

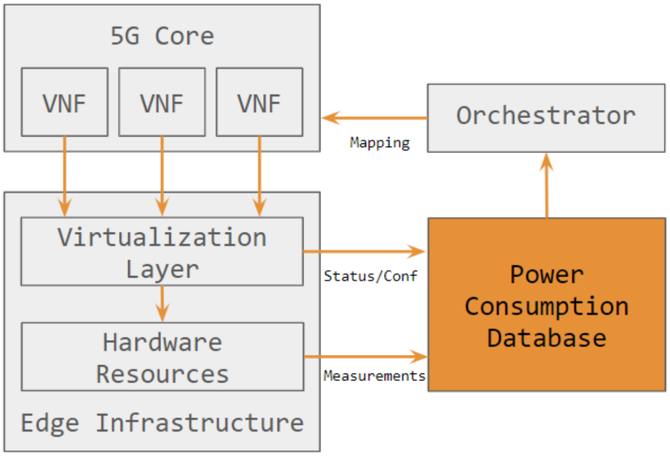

Fig. 1. Diagram showing a power consumption monitoring system integrated with a 5G NFV deployment. The parts we are concerned about in this work are highlighted in orange.

(OAM) decisions, enhancing it to support energy saving as one of the optimization parameters. The conceptual integration between this mechanism and a 5G NFV deployment is shown in Fig. 1. Power measurements are continuously gathered from the hardware infrastructure together with the status and configuration of the virtualization layer. Such information is processed and forwarded to the network orchestrator which acts on the 5GC configuration and deployment accordingly.

In this context, the main contributions of our work are as follows:

- We investigate the trend of power consumption of an open-source implementation of a 5GC instantiated on COTS hardware, analyzing different deployment options.
- We provide insights on the non-linear relationship between power consumption and processor utilization.
- We present a proof-of-concept for a real-time monitoring system for the power consumption of a 5GC, outlining how it can be used to orchestrate VNFs at the edge.

The rest of the paper is structured as follows. In Section II, we provide an overview on the related work, highlighting the novelty of our contribution in the field. Section III describes the utilized methodology for our empirical assessment, the testbed components, and their setup. The experimental results and their discussion follow in Section IV. Finally, in Section V we draw the conclusions of this preliminary study, forecasting the possible future work on this subject.

II. RELATED WORK

The topic of measuring and modelling of power consumption in virtualized environment is not a new one and some heterogeneous works already exists in the literature. The survey in [5] identify seven dimensions of variability in this research space and proceed to give an exhaustive overview of the different challenges and approaches present in the current state-of-the-art.

Different performance comparison between alternative virtualization technologies are presented in [6]–[10].

The analysis in [6] is based on the benchmarking of independently executed instances of an Apache HyperText Transport Protocol (HTTP) server and of a Redis data structure store. The KPIs chosen to evaluate both services were the amount of storage required to host the application, the memory utilization and the CPU utilization (both in the idle state and with the running service). In addition, Redis was used to perform a latency analysis of different traffic scenarios. The work done in [6] was expanded in [7] using a similar setup and methodology but with the inclusion of a fourth virtualization technology called Kata containers which consists of a hybrid model combining containers and VMs.

The study presented in [8] conducts extensive experiments on the power and energy consumption of four of the most adopted hypervisors and a container engine. Six different hardware platforms have been considered, including different rack servers architectures, one desktop server, and one laptop. The power measurements are gathered during a variety of computation-intensive, memory-intensive, and mixed Web server-database workloads at different level of intensity. The results highlight the different characteristics of each hypervisor and their aptness for specific workloads or platform with no one strictly outperforming the others.

In [9] the authors present the results of an empirical investigation comparing KVM, Xen, Docker and LXC: four virtualization technologies, the first two based on hypervisors and the last two on containers. The aim of the study is to characterize the power consumption of the considered virtualization technologies in the idle state and under CPU, memory and networking intensive workloads. The results show a significant difference between the case studies only during the network performance analysis. Container environment, and Docker in particular, register a lower power consumption compared to the hypervisor environments with Docker performing the best and Xen the worst.

In [10] a similar study is presented, comparing the overhead of virtualization technologies while performing network tasks. Results show that both hardware virtualization and paravirtualization solutions may consume 40% more energy and 5 time more CPU cycles than a standard bare-metal machine performing the same tasks. On the other hand, a container virtualization system, such as OpenVZ, is much more energy efficient, with a consumption comparable to the bare-metal machine.

While multiple studies have analyzed the performance and power consumption overhead of virtualization for the generic datacenter and cloud computing environment, "[...] the impact of virtualization technologies on power consumption in public telecommunication networks (PTNs) is still unclear" [5].

An analysis of NFV technologies applied to the Evolved Packet Core is presented in [11], comparing the environmental sustainability and energy requirements of a virtualized serving gateway to the business-as-usual solutions available on the market.

Greater attention has been directed towards characterizing and improving the energy efficiency of the RAN. In [12], the authors present an extensive data collection campaign including traffic, energy saving and power consumption statistics of a 5G active antenna unit. These metrics are then used to develop a novel artificial neural network architecture for modeling and estimating the power consumption of multiple 5G base station products.

The study presented in [13] examines the impact of virtualization on the RAN and in particular it compares the power consumption of the Base Band Unit (BBU) deployed using commercial devices and pools of VNFs. The examination is based on the analysis of publicly available traffic datasets and datasheets for both the commercial devices and the VNF pools. The results show that the exclusive use of virtual BBU can increase the power consumption up to 250% compared to commercial units, however a deployment mixing both solutions can achieve 20% energy saving while maintaining most of the advantages of VNFs.

As opposed to previous efforts or in addition to them, our work focuses specifically on the evaluation of the power consumption associated to different deployment approaches, including virtualized ones, of 5G mobile core networks. As for the virtualization domain, we investigate the overhead of VM-based deployments and container-based ones, as they are by far the most widespread frameworks.

III. EVALUATION METHODOLOGY AND TESTBED SETUP

All the results that will be presented later are gathered from an experimental testbed that we developed. As shown in Fig. 2, it comprises two layers:

1) the *communication layer*, which consists of the necessary hardware and software elements to build a complete, end-to-end 5GS;
2) the *power monitoring layer*, which consists of the necessary hardware and software elements to monitor the power consumption of the communication layer.

A. Communication Layer Setup

The communication layer represents the hardware used for hosting network processes and functionalities as well as the corresponding software components (e.g. VMs, containers, etc.). To simulate the server infrastructure commonly available at the network edge, we built a proof-of-concept testbed considering a simple setup and COTS hardware. Such testbed comprises five Intel® NUC units forming the 5GS. Three NUCs share the same hardware specifications, that are an Intel® i5-7260U processor and 8GB of RAM, as well as the same version of Ubuntu 20.04 Desktop operating system (OS). This is a requirement in order to obtain comparable results not biased by the hardware. On each of these pieces of COTS hardware, a fully functional instance of 5GC implemented by means of the Open5GS project[1] is deployed, according to three different deployment options:

1) *bare metal* deployment: the 5GC is installed directly on the OS;
2) *hypervisor-based* deployment: the 5GC is installed inside a VM running a Ubuntu 20.04 Server OS on top of a QEMU-KVM hypervisor[2];
3) *container-engine-based* deployment: 5GC is deployed inside a Docker environment with each network function running in a different container based on the *docker_open5gs* project[3].

As a general remark, let us observe that *containerization* has been gaining a lot of traction in recent years to the detriment of the more traditional hypervisor-based virtualization approach. The reason for this trend is that containers provide a much lighter and agile virtual environment by sharing kernel with the host OS while still providing a level of process and resource isolation. Thus, they present a great alternative, especially for highly dynamic VNF deployments.

The three types of deployments are represented in Fig. 3 with a schematic version of their system architecture. These deployments provide a small example of all possible virtualization solutions and were selected since they are the most commonly adopted and supported. The other two NUCs have an Intel® i5-1145G7 processor and 8GB of RAM running Ubuntu 20.04 Desktop OS. On one of these two, we installed UERANSIM[4], an open-source RAN and User Equipment (UE) simulation software that can connect to the three above mentioned 5GC instances. The fifth and final NUC serves as an endpoint for the UE traffic in the Data Network (DN).

B. Power Consumption Layer Setup

The power consumption layer represents the software and hardware that we propose to integrate in the network infrastructure in order to monitor the power consumption. In the case of the testbed used in this paper, each of the three NUC running the 5GC instances is connected to a Meross MSS310[5] smart plug capable of monitoring its power consumption in real-time. Thanks to open-source community projects which provide interesting insights about the Meross protocol[6][7], we were able to put in place a metrics retrieval system by which the smart plugs can be queried locally using specifically crafted HTTP packets. This allows us to collect the data without the need of connecting to the remote Meross servers, overcoming the stringent limitations on the requests rate.

Other metrics regarding the utilization of hardware resources, such as CPU and memory, can be gathered directly from the NUCs. Finally all metrics are collected and stored in a central location using Redis[8] both as a database and message broker.

---

[1] https://github.com/open5gs/open5gs. Last visited: January 2, 2024.

[2] https://www.qemu.org/. Last visited: January 2, 2024.

[3] https://github.com/herlesupreeth/docker_open5gs. Last visited: January 2, 2024.

[4] https://github.com/aligungr/UERANSIM. Last visited: January 2, 2024.

[5] https://www.meross.com/all/smart-wi-fi-plug-with-energy-monitor/6. Last visited: January 2, 2024.

[6] https://github.com/krahabb/meross_lan. Last visited: January 2, 2024.

[7] https://github.com/arandall/meross. Last visited: January 2, 2024.

[8] https://redis.io

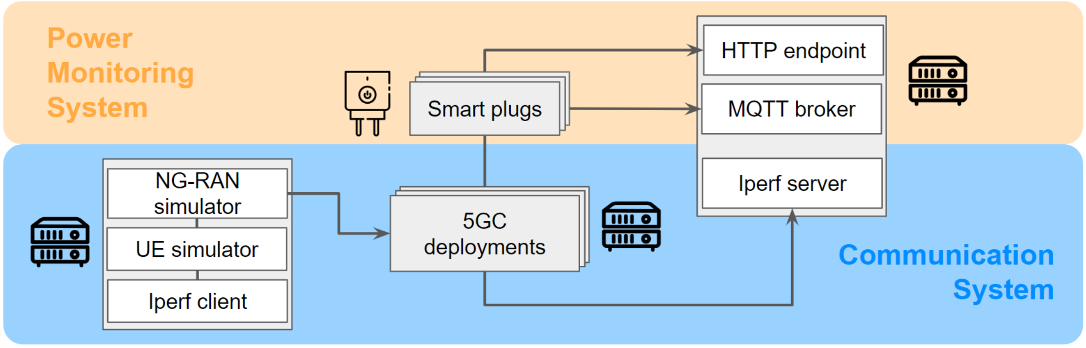

Fig. 2. Testbed architecture. We distinguish between two layers, namely the power monitoring layer (top part of the figure, in yellow) and the 5G communication layer (bottom part of the figure, in blue).

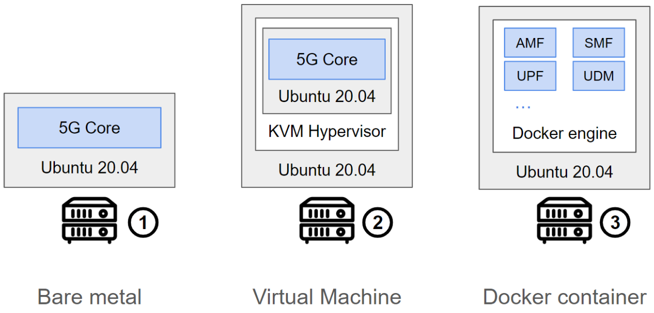

Fig. 3. Software architecture of the three virtualized 5GC deployments that will be analyzed in this study.

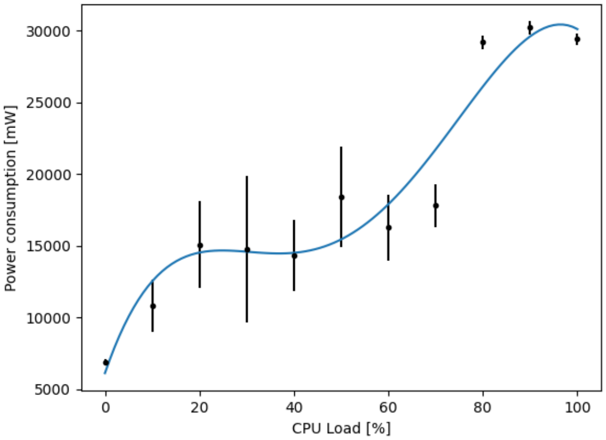

Fig. 4. Power consumption (in mW) against CPU load (in percentage with respect to the maximum achievable load) measured on the testbed hardware. We interpolated the data with a fourth degree polynomial function showing the nonlinearity.

IV. EXPERIMENTAL RESULTS

A. *Characterization of Hardware's Power Consumption*

Our first goal was to characterize the specific relationship between the CPU load and the power consumption in the Intel® NUCs. We remark that the value of the CPU load is a good approximation of the overall CPU usage, even if it does not take into account the resources used by the OS and other background processes, which me measured being less than 1% of the total computing resources. In order to perform this measurement, we generate an artificial CPU load using the *stress-ng* tool available on the Ubuntu OS, with incremental steps of 10%.

An assumption that is typically made, both by intuition and in some models found in literature is that a *linear* dependence exists between these two parameters. However, from our own measurement campaign, we found out that this is definitely not true for the adopted hardware, as shown in Fig. 4. In particular, please note the steep increment in power consumption for CPU loads greater than or equal to 80%. This is due to the Intel® Turbo Boost mechanism, which intervenes increasing the processor base frequency from 2.20 GHz up to the maximum value of 3.40 GHz in order to keep up with the high processing demand.

B. *Characterization of 5GCs' Power Consumption*

The main results of our study are presented in Fig. 5, showing the different power consumption for the three 5GC deployments described in Sec. III-A while varying offered traffic loads. The power consumption values are the average of the measurements gathered by sampling 10 consecutive minutes of the records from the monitoring system every 10 seconds, for each scenario. The error bars represent the 95% confidence interval on the data, assuming a normal distribution. The initial measurements are gathered in an idle state, meaning that the 5GC is installed and running on the system but no traffic is passing through the network. All consecutive measurements are gathered with different levels of uplink traffic exchanged across the 5GC user plane. The traffic is generated using the *Iperf* tool bound to the virtual network interface created by UERANSIM and associated with a simulated UE, and is delivered to an Iperf server in the DN. The bandwidth provided by the underlying wired network to the NUCs is between 800 Mbps and 900 Mbps.

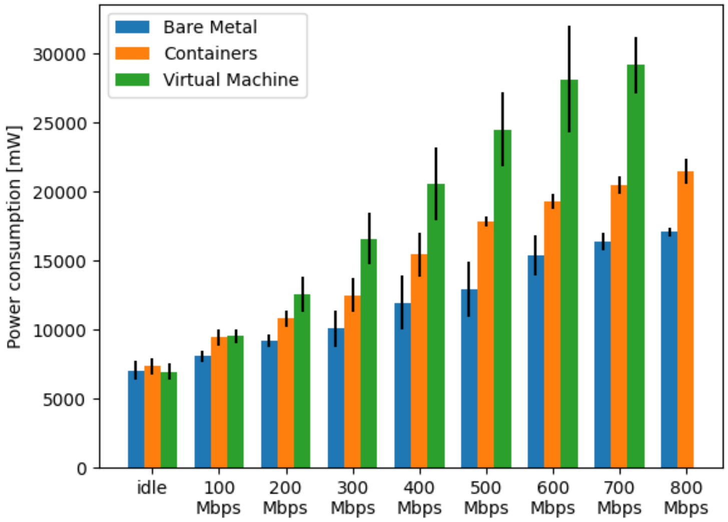

Fig. 5. Power consumption (in mW) for the three different virtualization technologies at different uplink traffic levels. Notice that the VM deployment cannot sustain a throughput higher than 700 Mbps, hence we do not show it past that point.

From the figure, it can be seen that the measurements in the idle states do not show significant difference in terms of power consumption among the three deployments. On the other hand, increasing the traffic throughput highlight a clear trend, yielding to the following observations.

- The highest throughput supported by all three deployments is of 700 Mbps. At that working point, the containerized and VM deployments require 25% and 78% more energy than the bare metal 5GC deployment, respectively.
- Both the bare metal and the Docker deployments can reach and overcome this working point. On the contrary, the deployment based on VMs cannot sustain a throughput higher than 700 Mbps, since the hardware reaches the 100% CPU usage, thus becoming the bottleneck of the communication system and therefore being unable to process anymore packets.

V. CONCLUSION AND FUTURE WORK

In a context where the importance of energy efficiency for networks is growing, this paper provided a preliminary assessment of the power consumption of a 5GC deployment based on three different virtualization options, that are, bare metal, virtual machines, and containers. Our analysis is based on a testbed we built using COTS hardware and open source software in order to resemble the infrastructure commonly available at the network edge. Different hardware architectures and software solutions are characterized by different power consumption patterns. The results of this preliminary study show how a real-time power monitoring system can be effectively paired with an existing 5GC deployment to gather power consumption metrics. The metrics can then be processed and forwarded to the orchestrator to enhance the energy efficiency of the network by means of green policies for the scaling, migration and consolidation of VNFs.

For our specific combination of software and hardware, we showed that a hypervisor-based 5GC deployment can be up to 78% more energy demanding than a similar bare metal deployment. On the other hand, a container-engine-based deployment, is only 25% more energy demanding while also providing a lighter and more flexible virtualization environment. These results are in line with the ones found in similar studies, confirming the validity of our implementation.

For future work on this topic, in the short-term, we plan to expand the testbed to include alternative hardware platforms and other open-source software. For example, other virtualization frameworks, such as unikernels, and different hypervisors or container engines may be investigated. Moreover, regarding the power monitoring system, the adopted solution proved to be good enough for the scope of our work, however a more accurate and reliable measuring equipment is required when dealing with smaller power variations and higher time resolution.

In the long-term, we plan to integrate this monitoring system in a wider network orchestration platform. This will allow the gathered metrics to be leveraged in the design and implementation of green policies and optimization algorithms for the OAM of both public and private 5GC networks.

REFERENCES

[1] Freitag C, Berners-Lee M, Widdicks K, Knowles B, Blair GS, Friday A. *"The real climate and transformative impact of ICT: a critique of estimates, trends, and regulations."* Patterns 2021;2:100340.

[2] Ericsson, "A holistic approach to address RAN energy efficiency," https://www.ericsson.com/en/blog/2021/12/a-holistic-approach-to-address-ran-energy-efficiency. Last visited: January 2, 2024.

[3] Delforge, P.; Whitney, J. Data Center Efficiency Assessment—Scaling up Energy Efficiency Across the Data Center Industry: Evaluating Key Drivers and Barriers. Natural Resources Defense Council, IP:14.08-A, August 2014. Available online: https://www.nrdc.org/sites/default/files/data-center-efficiency-assessment-IP.pdf. Last visited: January 2, 2024.

[4] ITU-T, *"Carbon data intensity for network energy performance monitoring"*, L.1333, Sept. 2022.

[5] E.-V. Depasquale, F. Davoli, and H. Rajput, "Dynamics of Research into Modeling the Power Consumption of Virtual Entities Used in the Telco Cloud," Sensors, vol. 23, no. 1, p. 255, Dec. 2022, doi: 10.3390/s23010255.

[6] R. Behravesh, E. Coronado and R. Riggio, *"Performance Evaluation on Virtualization Technologies for NFV Deployment in 5G Networks,"* 2019 IEEE Conference on Network Softwarization (NetSoft), 2019, pp. 24-29, doi: 10.1109/NETSOFT.2019.8806664.

[7] V. Aggarwal and B. Thangaraju, *"Performance Analysis of Virtualisation Technologies in NFV and Edge Deployments,"* 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), 2020, pp. 1-5, doi: 10.1109/CONECCT50063.2020.9198367.

[8] Congfeng Jiang et al., *"Energy efficiency comparison of hypervisors"* 2016 Seventh International Green and Sustainable Computing Conference (IGSC), 2016, pp. 1-8, doi: 10.1109/IGCC.2016.7892607.

[9] 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.

[10] R. Shea, H. Wang and J. Liu, "Power consumption of virtual machines with network transactions: Measurement and improvements," IEEE INFOCOM 2014 - IEEE Conference on Computer Communications, Toronto, ON, Canada, 2014, pp. 1051-1059, doi: 10.1109/INFOCOM.2014.6848035.

[11] R. Bolla, R. Bruschi, F. Davoli, C. Lombardo, J. F. Pajo and O. R. Sanchez, *"The dark side of network functions virtualization: A perspective on the technological sustainability,"* 2017 IEEE International Conference on Communications (ICC), 2017, pp. 1-7, doi: 10.1109/ICC.2017.7997129.

[12] N. Piovesan, D. López-Pérez, A. De Domenico, X. Geng, H. Bao and M. Debbah, *"Machine Learning and Analytical Power Consumption Models for 5G Base Stations,"* in IEEE Communications Magazine, vol. 60, no. 10, pp. 56-62, October 2022, doi: 10.1109/MCOM.001.2200023.

[13] R. Bolla, R. Bruschi, F. Davoli, C. Lombardo and J. F. Pajo, *"Debunking the "Green" NFV Myth: An Assessment of the Virtualization Sustainability in Radio Access Networks,"* 2020 6th IEEE Conference on Network Softwarization (NetSoft), 2020, pp. 180-184, doi: 10.1109/NetSoft48620.2020.9165481.

[14] Armenta-Cano, F.A., Tchernykh, A., Cortes-Mendoza, J.M. et al. *Min_c: Heterogeneous concentration policy for energy-aware scheduling of jobs with resource contention.* Program Comput Soft 43, 204–215 (2017). https://doi.org/10.1134/S0361768817030021

[15] Delforge, P.; Whitney, J. Data Center Efficiency Assessment—Scaling up Energy Efficiency Across the Data Center Industry:Evaluating Key Drivers and Barriers. Natural Resources Defense Council, IP:14.08-A, August 2014. Available online: https://www.nrdc.org/sites/default/files/data-center-efficiency-assessment-IP.pdf