Benchmarking Network Functions on competing hardware platforms - A survey

Rubens Figueiredo

June 27, 2022

1 Introduction

Following the expansion of network services and the diversification of users using these services is a trend of increasing energy consumption. Current estimates for total Information and Communication Technologies (ICT) contribution of Greenhouse Gas Emissions (GHG) is 1.4-2%, of which the larger contributor of energy consumed are the access networks (around 70% of total power consumption). By 2030, up to 23% of the world's electricity of CO_2emissions could be caused by the development of this particular industry [1].

Where data centre efficiency has generally improved, due to hyperscaling and the cloudification of networks, the distributed nature of access networks, where a large number of small scale data centres and central offices spread out over certain areas and drawing power form the grid makes for a hard target for optimisation. [2] In addition, the provided Quality of Service must also be an essential design goal for current and next generation networks.

In general, in order to neutralise or offset the GHG emissions, there is the need to improve access networks energy efficiency on three levels: equipment, sites and networks. Different strategies involve automatic sleep/ shutdown cycles, dynamic adjustment of channels, deployment of liquid cooling, among others. Furthermore, the use of renewable energy sources, such as solar or wind power, should be factored in as a sustainable alternative to reduce GHG emissions. Nevertheless, given the sporadic nature of renewable energy sources and uneven sources of energy in different countries could prove a challenge to completely shift the access network infrastructure from non-renewable sources [2] [1].

Network Function Virtualisation (NFV) and Programmable Dataplane (PDP) technologies could be used as enablers for adaptive resource consumption schemes and intelligent scaling mechanisms. NFV decouples the packet processing functions from specific hardware with software components, called Virtual Network Functions (VNF), which can be scaled on demand [1] according to the workload requirements. PDPs allow to offset (parts) of the NF into hardware, thus accelerating the function by running it on hardware and saving CPU cycles.

In this work we propose a survey of research work done in the domain of NFV and energy consumption. In particular, we are interested in the methods used to profile the performance of VNFs, in terms of throughput, latency, energy consumption, among other metrics, and which tradeoffs exist. Additionally, we investigate what is the reported impact of different PDP architecture on NF performance, and which solutions have been discussed to improve energy efficiency.

2 Background and motivation

2.1 Energy efficient networking

2.2 Programmable dataplanes and Hardware acceleration

2.3 Network Functions

3 Systematic Literature Review (SLR)

Purpose of the study, research questions. inclusion criteria, search strategies, quality assessment, data extraction strategies, synthesis and reporting

In order to identify where the knowledge frontier in Green Networking lies and to further drive empirical studies, it is necessary to summarise, analyse and synthesise the existing body of work in this area [?]. By planning the literature survey with techniques employed by other scientific fields, this process can be achieved with valid, reliable and repeatable methods, while minimising researcher bias and the possibility of drawing the wrong conclusions.

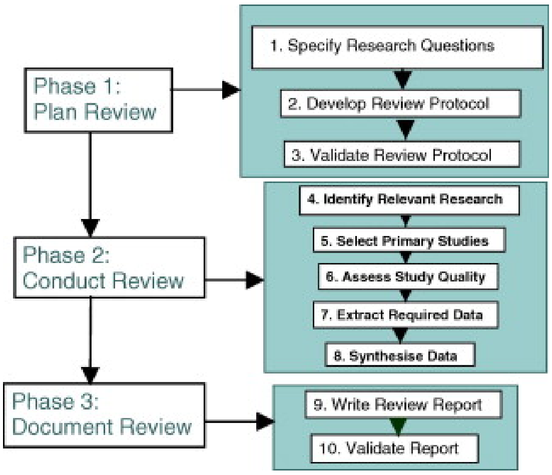

Figure 1: Systematic literature review process [3].

Not my figure, redraw and align with text

In this light, this survey aims to be a quasi-systematic review, where the research area is exhaustively described, without establishing a baseline comparison for the discussion of the results obtained [4].

The steps necessary for a successful Systematic Literature Review are summarised in figure 1. In short, this process consists of three phases: plan, conduct and document. In the planning phase, the objective and methodology of the study is laid out. This phase involves formulating the research problem to be answered, and identifying the proper methods that will help answer the research questions. In the conducting phase, the search strategy developed in the planning phase is executed, where the different sources are searched with relevant keywords, articles are screened, data is extracted and finally synthesised. In the documenting phase, the systematic review can be documented and the results gathered for the final document. In this section, we present the review protocol we followed for the development of this survey.

3.1 Planning Phase

3.1.1 Research questions

The research questions are essential to set a baseline for the following processes [**?**] in a SLR. The objective of this review is to identify the techniques used in performance profiling of NFs, with an emphasis on characterising the energy efficiency of systems being accelerated by different targets. As such, we have identified the following research questions (RQ):

RQ1 What are the different metrics employed to benchmark the performance of NFs?

RQ2 What are the strategies used to profile the performance of NFs?

RQ3 What is the impact of alternative hardware accelerators in the performance of NFs?

RQ4 What are the strategies (placement, orchestration, ...) to improve energy efficiency/ resource utilisation for hardware accelerated NFs?

3.1.2 Inclusion and exclusion criteria

This review considers all papers that meet at least one of the following inclusion criteria (IC) and does not satisfy any of the exclusion criteria (EC).

IC1 Study discusses the energy consumption measurements and benchmarks of NFs;

IC2 Study presents strategies for measuring the performance of NFs;

IC3 Study discusses hardware acceleration of NFs;

IC4 Study discusses strategies for improving the energy consumption/ resource utilization of NF;

EC1 Study not written in English;

EC2 Study has not been published in the 2017-2022 period;

EC3 Study dealing with topics irrelevant for the research questions;

EC4 Study not able to provide an adequate amount of information;

EC5 Study does not focus on the Network Functions, wired and wireless, and services thereof;

EC6 Study is not a primary study;

3.1.3 Search strategy

The field of NFV has been the focus of a lot of research in the past years and there is a lot of literature available. In order to reduce the amount of noise and to keep the focus of the research, a search strategy must be developed that fits the research questions. Since not one single source of data is able to find all of the relevant research work, multiple databases need to be searched [3]. The included databases are summarised in table 1.

ACM Digital library
Google Scholar
IEEEXplore
Springer
ScienceDirect

Table 1: Databases employed in the procedure of the review.

The keywords utilised for the development of the searches are summarised in table 2. The approach used to build the search strings is to include Subject, Technique and Objective. This approach is similar to PICO [4].

Subject	'Network Function', 'Network Function acceleration', 'Virtual Function', 'network function offloading', 'SFC', 'NF hardware acceleration'
Technique	benchmarking, profile, measurement, metrics, performance
Objective	'energy efficiency', energy consumption, 'power consumption', 'power efficiency'

Table 2: Search keywords based on 'classification'.

Pulled this classification out of thin air, need to find similar approaches for keyword aggregation.

After the initial phase of papers gathered, the papers are catalogued to the proper inclusion/ exclusion criteria and the research question they best fit. Then a backward and forward search is executed on the gathered papers. Backward searches involve searching for the cited papers, and forward searches involve getting the papers that have cited from the search results.

3.1.4 Selection process and quality assessment

Link to excel: https://docs.google.com/spreadsheets/d/1Z-3LYdBEAFCaiQ-71w2MO8Nco4GlbnQyTN7CL-r-MPs/edit?usp=sharing

Title	DB	Subtopics	Performance Metrics	Objective	Strengths	Weaknesses	IEC	RQ

Table 3: Excel classification of the gathered studies

The initial round of article search was done in April 2022. An excel document was produced where all the documents where sorted, as per the categories in table 3. When searching the databases, the search results are read through and the titles, abstracts and conclusions are read [4]. After the three sections of the article are read and if the information presented in the work fits the scope of the research questions, the research work is added to the excel database. In case of unclear information, then the article is included.

Due to non critical nature of this survey, no explicit quality assessment criteria as applied and as such, both journal and conference articles are included in the selection.

3.1.5 Information Extraction

The information extraction stage is central for the properly addressing the research questions. For RQ1-RQ3, the information extracted are the considered performance metrics, which can be obtained from the 'Evaluation' sections present in most considered research works. As for RQ4, the information extraction process is done with recourse to the 'System model' section, since the goal is to understand which methods have been proposed to improve energy efficiency.

3.2 Execution phase

After the planning phase, the execution of the search process takes place following the protocol presented in the previous section. Figure 2 shows the distribution of selected papers per-database [3].

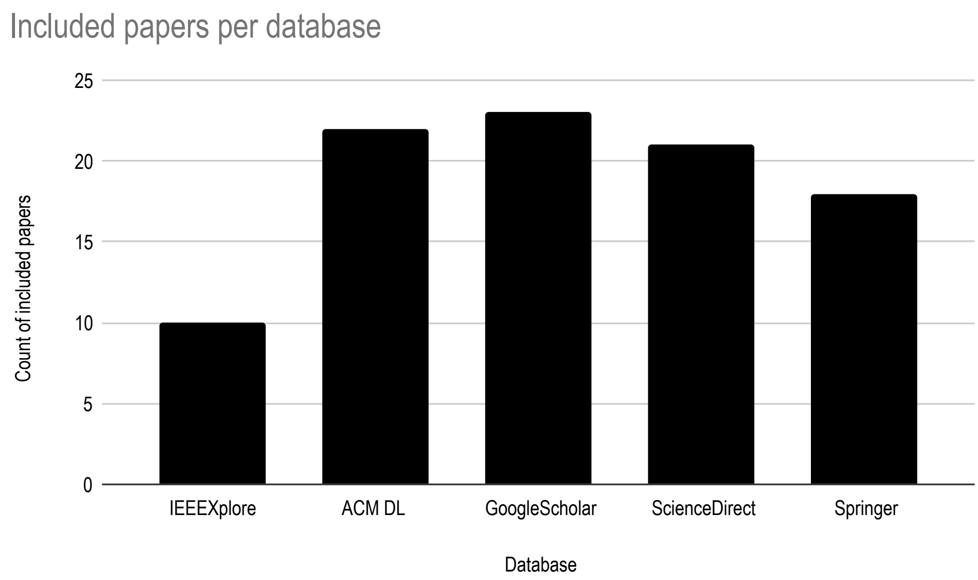

Figure 2: Count of included papers by database

Figure 3 displays the results of the search process in terms of the distribution of papers obtained per research question.

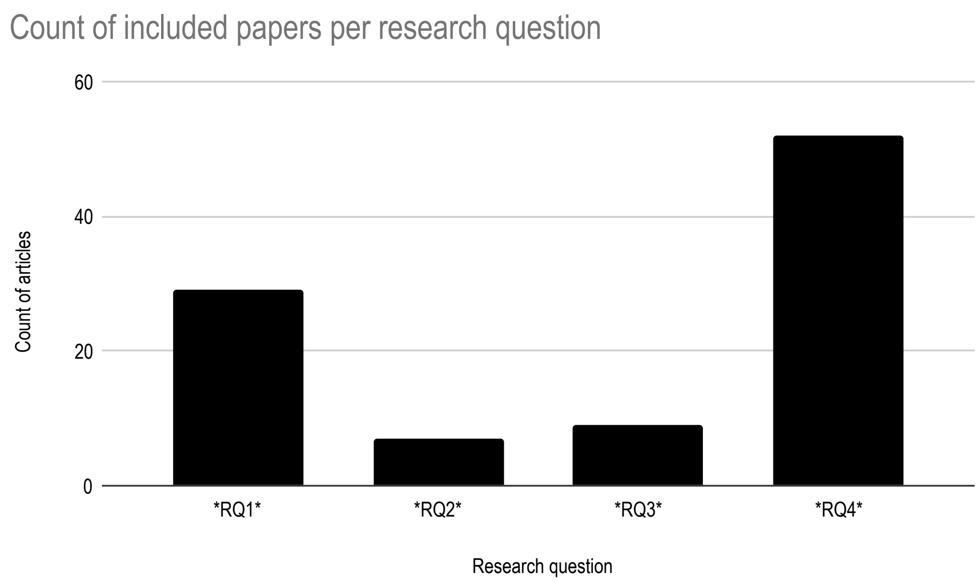

Figure 3: Count of included papers by research questions

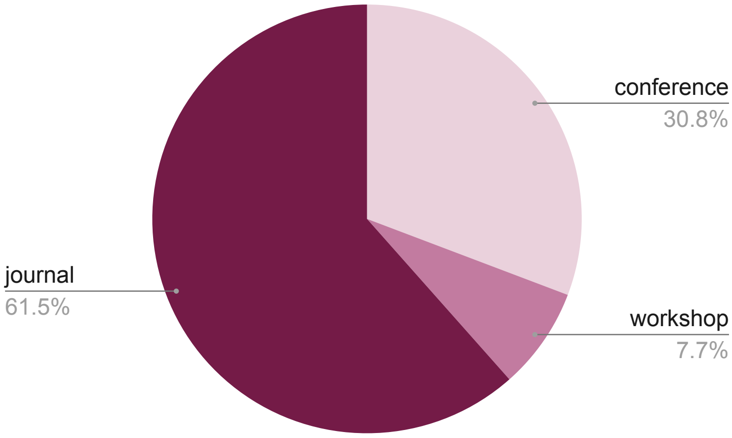

Figure 4: Distribution of included papers per distribution method

Figure 4 features the split of papers per distribution method.

4 RQ1: What are the different metrics employed to benchmark the performance of NFs?

HWA heterogeneity can contribute to more efficient ways of deploying NFs, however, the different underlying architectures can makes for a complex case of performance tuning. Operators must know the expected NF performance to scale up or down their services, without opening new attack vectors for service interruption [5]. Developers must know the underlying hardware characteristics to optimise their dataplane applications, and not always does the developer and the operator care for the same metrics/ know of the same metrics. [5].

Therefore, in this section we explore the benchmarking side of performance analysis of network function acceleration with PDPs and the benchmarks typically run to obtain an in-depth view of the systems performance.

5 RQ2: What are the strategies used to profile the performance of NFs?

Prediction, where techniques like static code analysis are done *a priori* to obtain the expected performance for certain NFs/ HWA combinations.

6 RQ3: What is the impact of alternative hardware accelerators in the performance of NFs?

Different types of Hardware accelerator, fpga, asic, smartnic.

discuss the research works that deal with NF HWA, and how to deploy them in order to

Shared resources: Heterogeneous packet processing frameworks do not handle well SFC since aggregated processing overhead and co existence interference overheads exist. [6]

For example, using hardware-acceleration to offload iFFT/FFT in cloud-RAN (C-RAN) scenarios to FPGAs, GPUs or DSP can result in power saving by about 70% per carrier [7].

Two popular modes of using hardware accelerators in NFV, look-aside and bump-in-the-wire. While look-aside acceleration is usually leveraged to offload compute-intensive operations; bump-in-the-wire acceleration processes the packets on the fly [7].

Stateful vs Stateless VNF offloading?

7 RQ4: What are the strategies (placement, orchestration, ...) to improve energy efficiency/ resource utilisation for hardware accelerated NFs?

Discuss ILP programs that *specifically* deal with energy consumption optimisations. While there exist models that optimise for resource utilisation in general, I would be interested in listing the algorithms/ heuristics/ meta-heuristics used to build models to improve EE. We can utilise the power consumption, the throughput/ delay/ jitter tradeoff, constraints, to find the different approaches that answer this research question.

A Network Service (NS) is composed of a number of NFs in a chain and is implemented under the interactions of these NF chains. Minimising the energy consumption of the NS involves the orchestration of the NF chains to minimise resource consumption and enhance reliability.

NOT MINE: The total energy consumption of a NFV-enabled telecom network consists of two major parts. The first part is the energy consumption of servers hosting VNF instances. The second part is the energy consumption of physical links transmitting traffic. For the energy consumption of servers, the authors of [11] note that though CPU is the predominant power consumer of a server, main memory, data storage, and networking devices also consume a considerable amount of energy that cannot be disregarded. Thus, the energy consumption of a server consists of two main parts: (1) the energy consumption for keeping itself on, and (2) the energy consumption produced by processing SFC requests, which is positively related to CPU utilization [8].

If a server is on but is processing no load, then it is consuming idle power. When load is offered to the server, it consumes the idle power as well as the power necessary to operate the systems servicing the load. If it is off, it consumes no power. The main mechanism for saving energy in cloud computing is to consolidate the workload on as few servers as possible and putting the others in lower power mode or shutting them down [9].

MEC and VNF consolidated architectures may prove unbalanced, as the MEC application latency requirements, however many algorithms designed to optimise for energy efficiency may not scale well for other demands [10].

Research work	Heuristic/ Optimisation method	Energy saving method	
[9]	ILP	VNF consolidation	Flow secu
[7]	ILP/ VNF-AAPC	Minimise servers	Hard
[10]	(Fuzzy) Formal Concept Analysis	Minimise servers	co
[11]	Reinforcement learning (A2C, PPO)	minimise the amount of redundant VNFs	
[12]	Student Placement Algorithm	Offloading decision at different levels	
[13]	HEAP	Orchestrate IoTNFs and VNF in a energy aware way	
[8]	MC-EEVP	Place NF to minimise server and link power	C

Table 4: Comparison between different proposed models and solving heuristics

8 Discussion

A challenge is how to collect, store, and retrieve the information which is the input for our RL agent [11].

9 Conclusion

References

[1] C. Tipantuña and X. Hesselbach, "NFV-Enabled Efficient Renewable and Non-Renewable Energy Management: Requirements and Algorithms," *Future Internet*, vol. 12, p. 171, Oct. 2020.

[2] "Green Future Networks: Sustainability Challenges and Initiatives in Mobile Networks," p. 37.

[3] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *Journal of Systems and Software*, vol. 80, pp. 571–583, Apr. 2007.

[4] A. M. Magdaleno, C. M. L. Werner, and R. M. d. Araujo, "Reconciling software development models: A quasi-systematic review," *Journal of Systems and Software*, vol. 85, pp. 351–369, Feb. 2012.

[5] R. Iyer, K. Argyraki, and G. Candea, "Performance Interfaces for Network Functions," p. 19, 2022.

[6] Y. Hu and T. Li, "Enabling Efficient Network Service Function Chain Deployment on Heterogeneous Server Platform," in *2018 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, pp. 27–39, Feb. 2018. ISSN: 2378-203X.

[7] G. P. Sharma, W. Tavernier, D. Colle, and M. Pickavet, "VNF-AAPC: Accelerator-aware VNF placement and chaining," *Computer Networks*, vol. 177, p. 107329, Aug. 2020.

[8] Z. Xu, X. Zhang, S. Yu, and J. Zhang, "Energy-Efficient Virtual Network Function Placement in Telecom Networks," in *2018 IEEE International Conference on Communications (ICC)*, pp. 1–7, May 2018. ISSN: 1938-1883.

[9] S. Demirci, M. Demirci, and S. Sagiroglu, "Optimal Placement of Virtual Security Functions to Minimize Energy Consumption," in *2018 International Symposium on Networks, Computers and Communications (ISNCC)*, (Rome), pp. 1–6, IEEE, June 2018.

[10] W. Khemili, J. E. Hajlaoui, and M. N. Omri, "Energy Aware Fuzzy Approach for VNF Placement and Consolidation in Cloud Data Centers," *Journal of Network and Systems Management*, vol. 30, no. 3, pp. 1–29, 2022. Publisher: Springer.

[11] G. L. Santos, T. Lynn, J. Kelner, and P. T. Endo, "Availability-aware and energy-aware dynamic SFC placement using reinforcement learning," *The Journal of Supercomputing*, vol. 77, pp. 12711–12740, Nov. 2021.

[12] L. Zhang, "Energy-delay tradeoff for virtual machine placement in virtualized multi-access edge computing: a two-sided matching approach," p. 19.

[13] R. Kouah, A. Alleg, A. Laraba, and T. Ahmed, "Energy-Aware Placement for IoT-Service Function Chain," in *2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pp. 1–7, Sept. 2018. ISSN: 2378-4873.