Demonstration of Energy Efficient Optimization in Beyond 5G Systems supported by Optical Transport Networks

P. Georgiadis(1), M. Anastasopoulos(1), A.-I. Manolopoulos(1), V.-M. Alevizaki(1), N. Nikaein(2), A. Tzanakaki(1)
(1) National and Kapodistrian University of Athens, Greece (2) Department of Communication Systems, EURECOM, France

Abstract: The paper proposes an optimization framework using artificial intelligence to optimize the energy efficiency of a B5G system operating over an SDN controlled optical transport network. The system is evaluated over an operational B5G testbed.

OCIS codes: (060.4254) Networks, combinatorial network design; (060.4256) Networks, network optimization

1. Introduction

6G networks are expected to provide services for a wide spectrum of vertical applications with greatly varying requirements in terms of capacity, latency, reach and mobility with minimum environmental footprint. To achieve this, complex 6G infrastructures, involving an excessive volume of operational parameters, need to minimize their energy consumption. This can only be performed through the adoption of flexible management and control frameworks relying on advanced Artificial Intelligence (AI) schemes able to support optimized service delivery through automated decisions enabled by a holistic self-learning approach. To facilitate this, 6G platforms need to collect and process large volumes of data, ensuring high level of data privacy, security and trust. This vision emphasizes the need of a transport network that offers increased connectivity capabilities and features to support the required data collection and management as well as the adoption of AI techniques. It is clear that optical networks, given their increased capacity and flexibility, can offer the most realistic technical approach to realize this vision.

Towards this direction, this study considers a multi-vendor optoelectronic transport network able to handle both fronthaul and backhaul network traffic supporting at the same time a wide range of connectivity options (point-to-point, point to multi-point and multi-point to multipoint) with capacities ranging between 1Gbps up to 100Gbps. In this context, a challenging issue that needs to be addressed is that of energy efficient service delivery across multiple network layers involving mechanisms operating at different time scales [1]. Traditionally, energy efficiency in converged optical-wireless networks supporting 5G services has been addressed using a variety of optimization algorithms. These include identification of optimal routing paths, baseband processing functional split options, adjustments in Radio Access Network (RAN) configurations, traffic offloading and multi-layer networking [2] taking appropriate decisions that minimize overall energy consumption. Therefore, energy efficiency can be achieved considering multiple aspects such as spatial and temporal traffic statistics at the access network, optical transport node capabilities, QoS requirements, background traffic in the network etc [3]. This in practice is performed through the optimization of a predetermined cost function with a known mathematical representation subject to a set of constraints. In an operational environment, this is considered by the infrastructure orchestrator to identify and execute optimal service delivery policies for improved energy efficiency. However, in 6G systems expected to operate in a fully disaggregated manner, identification of accurate cost functions that can be used by the orchestrator to achieve maximum energy efficiency is not a trivial task. Uncertainty in any of these functions (either objectives or constraints) may lead to erroneous decisions which can seriously deteriorate network performance.

To address this problem, Machine Learning (ML) techniques can be used to drive service delivery and the associated resource allocation decisions by leveraging data collected through the monitoring system. ML algorithms can be used to predict the system status in the upcoming time periods and these predictions can be exploited by the orchestrator to take improved/informed decisions regarding optimal resource allocation for future service requests. Usually, the orchestrator calculates the optimal decisions/policies applying reinforcement learning techniques [4]. However, a main disadvantage of these models is that they require several iterations to converge, while in many cases these iterations are very expensive and even destructive for the system. At the same time, the ML model predictions require large datasets to forecast parameters with sufficient accuracy levels. To combat these limitations, we propose to combine a specific class of ML models referred to as Gaussian Process Regression (GPR) requiring limited input data, with Bayesian Optimization to identify optimal energy efficient strategies in 6G networks [5]. The evaluation of the proposed scheme is performed over an experimental Beyond 5G (B5G) testbed interconnected through a multi-vendor Software Defined Networking (SDN) controlled optoelectronic transport network.

2. Overall System Architecture

We consider a B5G system comprising a set of multi-vendor SDN-controlled optoelectronic switches with different capabilities (in terms of number of ports, capacity per port, latency) interconnecting Remote Units (RUs) with compute

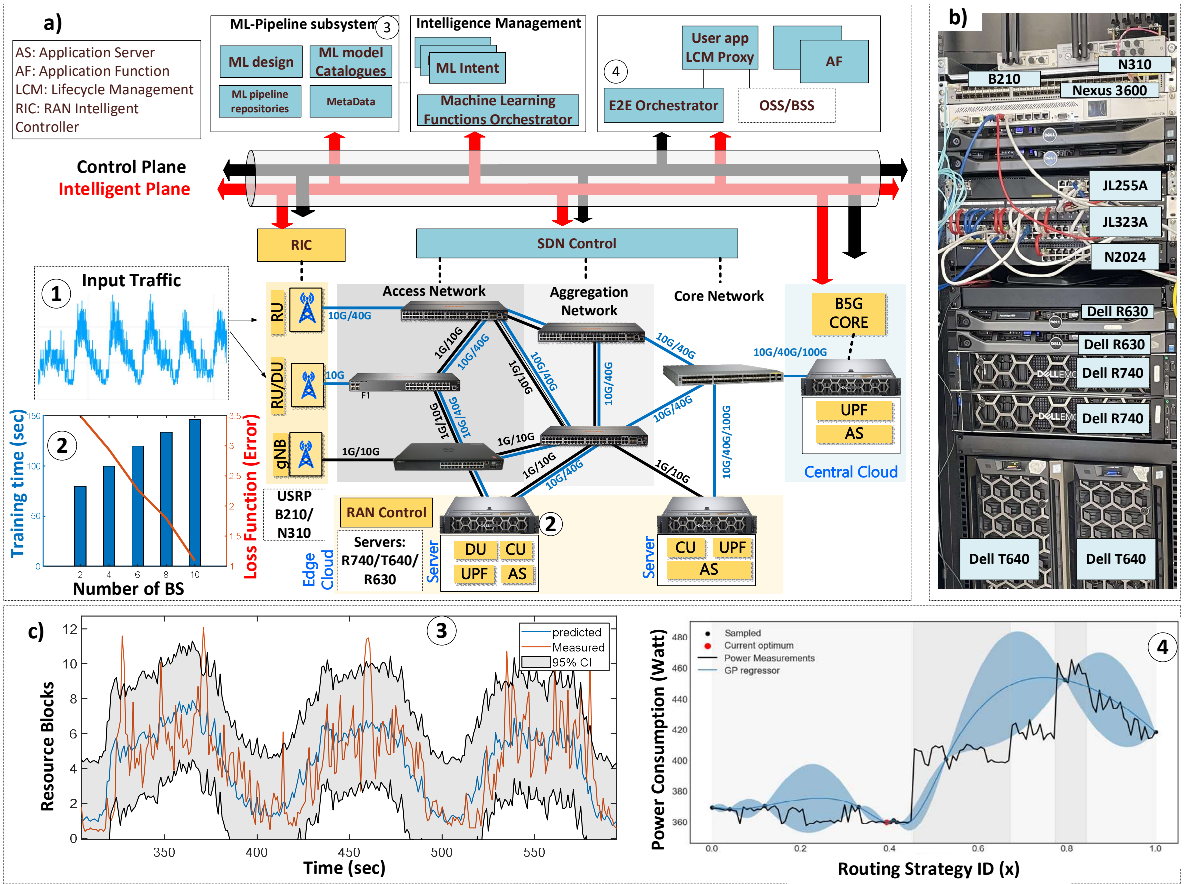


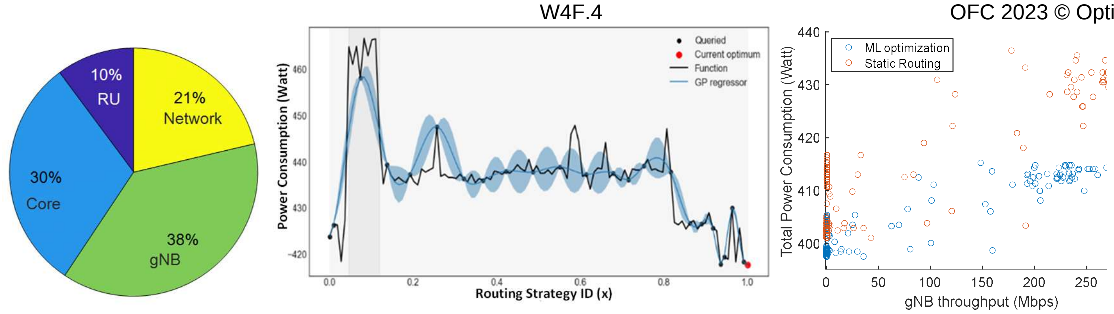
Fig. 1a) Network configuration, b) Experimentation infrastructure, c) ML pipeline for energy minimization: 1) Input traffic flows, 2) Optimal Placement of ML functions, 3) GPR model estimating RBs in RAN, 4) E2E Energy-aware optimization: The Bayesian solver identifies optimal routing strategy (red dot) minimizing power consumption. Results extracted for incoming traffic 150Mbps.

nodes hosting 5G data plane functionalities implemented based on OAI (https://openairinterface.org/). The opto-electronic transport network is organized in a hierarchical manner (Fig.1 a) offering RU connectivity, collecting and aggregating transport traffic from various cells to a central location. The access network is equipped with low energy consuming switching nodes having limited number of input ports and relatively small capacity (capacity 1/10GbE per port, with SFP, SFP+ and RJ 45 transceivers). High-end switches with higher capacity and density (10G/40G/100GbE/ports) are placed in the aggregation and core network segments offering much higher energy efficiency (lower bit/Joule) under high utilization compared to low-end switches. For the compute domain, we consider edge servers attached to the access switches and central cloud servers connected to the aggregation/core switches. Depending on the mobile service latency requirements, various functions can be placed at different network locations. For this system set-up, a critical decision is where and through which paths RU demands will be routed to the appropriate compute facilities.

As traffic demands have specific requirements in terms of isolation, security and QoS guarantees, appropriate management and end-to-end orchestration mechanisms play a key role. These mechanisms are supported by the top layer of the proposed architecture (Fig.1a) offering tools that allow design of new services, monitoring and management of services throughout their lifetime as well as service onboarding. In accordance with the Zero touch network & Service Management (ZSM) framework [6] and in order to address the highly complex environment of B5G networks, an intelligent/semantic management plane is introduced that interacts with all building blocks of the system enabling appropriate decision making for optimal system performance. This is achieved through a purposely developed intelligent management framework relying on a set of building blocks aligned with the ITU-T Y.3172 [7] standard providing a set of microservices including: i) design of custom ML pipelines, ii) orchestration of ML pipelines through the ML Function Orchestration (MLFO).

3. Proposed Model and Numerical Results

To minimize the power consumption an ML-pipeline [8] has been designed and deployed in the testbed of Fig.1. This pipeline performs a set of operations including: *1) collection of* energy, traffic, and resource utilization statistics. This functionality is achieved through appropriate interfaces implemented across the different layers of the platform. Monitoring data are transferred using a dedicated slice ("intelligent plane") as shown in Fig.1a) with 1sec sampling



a) b) c)

Fig. 2. a) Power consumption per domain, b) Energy-aware optimization using Bayesian techniques: Example for traffic/gNB 300Mbps, c) Total power consumption as a function of traffic/gNB for the static and dynamic routing using ML optimization.

frequency. The relevant metrics are stored in a distributed manner either at the edge or at the central cloud facility. The decision where each collected measurement is stored is taken by the MLFO after validating the tradeoffs between accuracy and complexity. An example is shown in Fig.1 a) (2) where the tradeoff between prediction accuracy (measured in terms of Loss-function) and complexity (training time) for different number of base stations (BS) storing their statistics in a collocated edge server is depicted. Numerical evaluations have been carried out using a GPR model predicting mobile network traffic with optimized hyperparameters. *2) Predictive Analytics ML model training:* In the second step of the ML pipeline, descriptive and predictive statistics of the 5G platform affecting the provisioned service are extracted. Typical examples are shown in Fig.1c) where two GPR models have been trained predicting the resource blocks (RB) of a specific BS in the RAN and the waiting delay in the CPU of the containerized DU/CU functions. The relevant trained ML models are stored in the ML-repository which can be then used by the end-to-end orchestrator as constraints. 3) *Prescriptive analytics:* In the next step of the ML pipeline, an ML model that optimizes the system for energy efficiency is created. This is achieved applying Bayesian Optimization techniques with the objective to identify optimal strategies (optical transport routing paths) minimizing energy consumption. The relevant problem can be written as a constraint Bayesian Optimization problem of the form [5]: $min\,\mathcal{E}(\mathbf{x})\,, s.t.\ \mathrm{c}(\mathbf{x}) \le \Lambda\,.$ where $\mathcal{E}(\mathbf{x})$, is the energy consumption function under routing strategy $\mathbf{x}$ and $\mathrm{c}(\mathbf{x}) = [c_1(\mathbf{x}), \dots, c_m(\mathbf{x})]$, $\Lambda = [\lambda_1, .., \lambda_m]$ is the set of inequality constraints that need to be satisfied for the system to operate. These functions are estimated using the corresponding GPR models available through the previous step of the ML pipeline. The Bayesian optimizer available through the ML repository solves the problem and identifies the optimal routing paths **x.** As shown in Fig.1 c (4), GPR is used to approximate $\mathcal{E}(\mathbf{x})$ with specific probability under different routing strategies $\mathbf{x}$ and the Bayesian solver finds routing policy $\mathbf{x}$ that minimizes power consumption. *4) Deployment*: In the final step, the decision is communicated to the SDN controller to apply optimal routing policies.

We evaluate the performance of the proposed system configuration using the testbed shown in Fig.1b). The power consumption of the system is monitored through metered outlet power distribution units. The input traffic was generated using Iperf. Fig.2a) shows the average power consumption per segment. The RU corresponds to the power consumption of the USRP N310. The network consumption accounts approximately for 21% of the total energy consumption. It should be noted that all measurements have been carried out using short reach transceivers. Fig.2b) shows a snapshot of the Bayesian optimization process for a scenario where the incoming mobile traffic per gNB is 300Mbps. Through sampling history measurements, the Bayesian solver approximates power consumption as a function of the available routing strategies and determines the specific routing policy (pointed out with the red dot) that minimizes the total power of the system. Finally, a comparison between the proposed policy and a preconfigured static routing policy is shown in Fig.2c). The proposed model setup tunnels between the RUs and the application server (AS) through the most energy efficient paths in the optical domain reducing by an average of 50% of the optical network power consumption and 10% of the total system consumption.

Conclusions

We proposed and experimentally evaluated an AI based optimization framework to minimize energy consumption in B5G system exploiting optical transport networks.

Acknowledgements

This work has been financially supported by the 5G PPP EU projects 5G-COMPLETE and 5G-VICTORI.

4. References

[1] 37.817 - Study on enhancement for data collection for NR and ENDC, https://portal.3gpp.org
[2] A. Tzanakaki et al., "Converged Optical, Wireless, and Data Center Network Infrastructures for 5G Services," JOCN, 11, 2019
[3] V. -M. Alevizaki, et al., "Distributed Service Provisioning for Disaggregated 6G Network Infrastructures," IEEE TNSM, 2022
[4] Z. Gao et.al., "Deep Reinforcement Learning for BBU Placement and Routing in C-RAN," OFC 2019, paper W2A.22.
[5] J.- R. Gardner et.al. "Bayesian optimization with inequality constraints". ICML'14). JMLR.org, II–937–II–945.
[6] ETSI GR ZSM 004 V2.1.1 (2022) Zero-touch network and Service Management (ZSM); Landscape/
[7] Y.3172 (06/19), Architectural framework for machine learning in future networks including IMT-2020
[8] ORAN-WG2.AIML.v01.00 Technical Report O-RAN Working Group 2 AI/ML workflow description and requirements