TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)
31 Oct - 3 Nov 2023. Chiang Mai, Thailand

Optimizing Funcitional Split in 5G Cloud RAN: A Particle Swarm Optimization Approach

Wai Phyo
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University, Bangkok, Thailand
6272083321@student.chula.ac.th

Pruk Sasithong
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University, Bangkok, Thailand
pruk.sasithong@gmail.com

Lunchakorn Wuttisittikulkij
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University, Bangkok, Thailand
Lunchakorn.W@chula.ac.th

Shashi Shah
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University, Bangkok, Thailand
Shashi.S@chula.ac.th

Abstract— **The Cloud Radio Access Network (C-RAN) is an innovative technology with great promise for minimizing wireless network deployment and maintenance costs. In this study, our main goal is to reduce the costs associated with functionally placing the RAN while accounting for the computational expense and the front-haul bandwidth usage among various users. To achieve this, we propose to apply Particle Swarm Optimization (PSO) to achieve effective allocation of computational resources and the front-haul bandwidth, ensuring an efficient and cost-effective C-RAN design. Experimental results on different traffic have shown that the proposed PSO can provide cost-effective design of the C-RAN as compared to the optimal solution from the integer linear programing (ILP) approach.**

Keywords— Cloud RAN, Particle Swarm Optimization, Integer Linear Programming, Functional Split

I. INTRODUCTION

The rapid growth of innovative services like Metaverse, Virtual Reality (VR), Augmented Reality (AR), self-driving cars, Internet of Things (IoT), cloud computing, and other cutting-edge technologies has been a significant factor in the exponential increase of mobile device traffic in recent years. Building a cost-efficient network that can guarantee high-quality service delivery and use various architectural technologies is essential to overcoming these obstacles. A cloud radio access network is an architecture for radio access networks that is centralized, based on cloud computing, and supports 2G, 3G, 4G, and future wireless communication protocols. This strategy aims to create a RAN design that is both economical and energy-efficient. C-RAN architecture enables us to simplify Base Stations that share signal capabilities with numerous antennas [1]. This suggestion aims to reduce costs by optimizing the C-RAN functional splits. Enhancing the performance of the quality of service (QoS) is one of the objectives of RAN optimization, and discovering the ideal combination of RAN parameters is another [2]. It can be achieved by increasing flexibility and lowering the cost of infrastructure deployment by implementing the functional split in RAN. To reduce costs and optimize bandwidth utilization, we will propose the Particle Swarm Optimization method in this paper.

This paper contributes to the field of cloud radio access networks by exploring the utilization of particle swarm optimization for optimizing bandwidth and processing costs. The primary objective is to minimize the RAN functional placement cost, which encompasses computational costs and fronthaul bandwidth utilization across multiple users. Furthermore, the study highlights the importance of fine-tuning the iteration parameter in the deployment processes to achieve desirable cost outcomes. By carefully adjusting the number of iterations, the deployment costs tend to stabilize and converge, providing insights into optimizing the PSO-based process for reducing the cost of bandwidth and computing in C-RAN.

II. RELATED WORK (LITERATURE REVIEW)

The authors in [3] provide an end-to-end system analysis considering overall costs and energy usage. They suggest a mixed integer programming (MIP) formulation for the problem and use the IBM CPLEX Optimizer to analyze the impact of consumer delay requirements on decision-making. However, their model's application is constrained by its singular user emphasis. The author in[4] provides a detailed explanation of the tele-traffic theory to analyze the allocation of resource periods and the resource allocation rate at the front-haul link. And then, based on this, the author explains in detail the purpose of saving energy and costs to realize a great deal for baseband processing sub-units when front-haul or baseband processing resources become further expensive for an operator. Moreover, it proves that user traffic has a high impact on segregation. The author formulates the issue using the tele-traffic technique and uses OPNET, a discrete event-based simulator.

In a paper [5], the authors present a technique for achieving energy efficiency in the 5G infrastructure that utilizes integer linear programming and an LSTM-based neural network. Their strategy focuses on maximizing the functional split of optical transport to reduce overall energy consumption. It is crucial to remember that the split's deployment is generally cell-centric, which could limit its application in situations calling for flexibility or user-centric split setups. The authors of [6] propose an ILP formulation of the problem and use the Lagrangian relaxation algorithm to minimize the number of routes that may lead to the release of critical services delay while also optimizing the cost of using baseband processing, which uses across multiple websites. Despite the network calculus technique, this function does not offer a quantitative standard method to calculate the cost of the requirements for each split. It is comprehensive to measure the delay on links on optical and wireless networks. The authors [7] focus on choosing functional divisions despite considering various types of cell interference. ILP and a heuristic approach solve the problem. They provide a novel heuristic technique to reduce the bandwidth used in the transport network and inter-cell interference. The functional split method performs at the cell level, which may have drawbacks in situations that call for a more fine-grained or user-centric approach.

In [8], the authors provide graph-based architecture, considering the advantages of the resulting path established by the front-haul connection and the latency requirements set by each of the cells to reduce the cost of computing the allocation of resources (RA) at two locations. Although the fundamental characteristics of natural systems reflect by the assumption of RA costs and latency constraints, The author formulates the problem using graph clustering and applies a genetic algorithm. The authors used a novel functional split orchestration scheme [9] to minimize the deployment cost of 5G C-RAN. They use ILP to generate the cost function for each split and PSO to optimize the cost function for each functional split. They proved their solutions had better performance for the resolution time and total deployment cost.

The study by the authors [10] presents a user-split orchestration approach that aims to reduce the front-haul link's energy and bandwidth consumption. The model uses quantitative models to calculate computing and link needs for each split. This strategy, however, is based on an inaccurate split model that treats the platform control function, including the MAC scheduler, as a user-centric processing unit. The IBM CPLEX optimizer handles the optimization work and the problem construct using Mixed Integer Programming (MIP). In the paper [11], a comprehensive model introduces to optimize the total cost of ownership (TCO) of a fiber-based RAN with split baseband processing units (BBU). The model takes quantifiable resource requirements for computation and links into account. Although it generates a split for each connected user in a cell, this coarse-grained approach might need to be revised. ILP is used to define the problem, and the IBM CPLEX optimizer is used to resolve it. For RAN optimization and control, the author in [12] proposed deep reinforcement learning based on the double Q network. It focused on choosing the suitable schedule configuration for each real eNodeB. They proved that the network performance outperformed the existing rule-based algorithm by applying deep reinforcement learning to an entire RAN system.

III. Problem Statement and Existing ILP Formulation

This section describes an existing ILP approach [11] for optimizing the C-RAN architecture's front-haul bandwidth usage and computational costs for Remote Radio Units (RRU) and Radio Cloud Center (RCC). The numerical evaluation outcomes are served as the benchmark against our proposed PSO methodology. In the problem, a set of K available splits made available to each of the total of N users to select from. Let x_i^k be a binary variable that is equal to 1, when user i is assigned with split k; otherwise, 0. We describe the RRU as a set of data units, each of which has a computational capacity of C_{RRU} GOPS, Power Consumption (P_U) and weight factor (α), and Power Usefulness Effectiveness (PUE_U). Additionally, we also describe the RCC as a set of data units, each of which is capable of doing computations with the following: computational capacity of C_{RCC} GOPS, Power Consumption (P_C), weight factor (β), and Power Usefulness Effectiveness (PUE_C). A_i^k (E_i^k) stands for the GOPS used at the RRU (or RCC) for the split k of user i, while is the front-haul bandwidth that the split k of user i generated. Front-haul link (FH) is used to connect RRU to RCC [9]. Thus, the problem can be formulated as follows: [11]

Minimize:

$$\alpha.\, PUE_U.\frac{P_{RRU}}{P_U} + \beta.\, PUE_C.\frac{P_{RCC}}{P_C} + \gamma.\frac{FH_{rate}}{B}$$

Subject to:

$$\sum_i^K x_i^k = 1 \quad \forall i \in N \qquad (1)$$

$$\sum_{i=1}^N \sum_{k=0}^K x_i^k R_i^k \le \mathrm{B} \qquad (2)$$

$$\sum_{i=1}^N \sum_{k=0}^K x_i^k A_i^k \le C_{RRU} \qquad (3)$$

$$\sum_{i=1}^N \sum_{k=0}^K x_i^k E_i^k \le C_{RCC} \qquad (4)$$

$$x_i^k \in [1]\,, \forall i \in N\,, \forall k \in K$$

Where:

$$P_{RRU} = (1/P_f).\sum_{i=1}^N \sum_{k=0}^K x_i^k A_i^k$$

$$P_{RCC} = (1/P_f).\sum_{i=1}^N \sum_{k=0}^K x_i^k E_i^k$$

$$FH_{rate} = \sum_{i=1}^N \sum_{k=0}^K x_i^k R_i^k$$

The constraint (1) states that each UE may only choose one split. The upper bound limit for the fronthaul link's total produced rate is expressed by constraints (2), (3), and (4), respectively, as well as the RRU and RCC computational resource requirements. After that, we determine the total amount of power used in RRU and RCC, respectively, P_{RRU} and P_{RCC}. FH_{rate} is used to evaluate the total generated rate of the Fronthaul link [16].

In a cellular network design, the baseband processing unit (BBU) and the radio unit (RU) are divided into different functional categories according to the 3rd Generation Partnership Project (3GPP) practical split idea. It outlines how the processing responsibilities are split between these two entities for a scalable and effective network operation. The concept of a functional split is especially applicable in the case of cloud radio access network (C-RAN) deployments, as the traditional baseband processing operations are split from the remote radio units and compiled in a BBU pool. This centralization enables enhanced coordination, better resource usage, and simpler network management. The 3GPP functional split offers a variety of alternatives for splitting the baseband processing tasks, allowing network operators and equipment suppliers to select the best configuration following their unique needs and deployment scenarios [2,6].In this paper, we assume that there are a total of seven split choices included in the 3GPP functional split, referred to as Splits 0 through 6. Within the Cloud Radio Access Network (C-RAN) architecture, each split choice denotes a certain arrangement of processing operations. The processing tasks carried out at the Remote Radio Head (RRH) level are included in Split 0. Processing tasks are no longer carried out at the RRH but are subsequently offloaded to the cloud or Baseband Unit (BBU) as we progress from Split 0 to Split 6. A more centralized and adaptable strategy can be used as a result of the split of processing tasks, with the computational workload being handled by the cloud or BBU. The numerous split choices enable the C-RAN architecture to adapt to varied deployment environments and optimize utilization of resources [2,6].

Fig. 1(a) and (b) show the results of computational cost and front-haul bandwidth utilization. The users are divided into different functional groups to make assigning users to specific amenities easier, which are identified by split options from 0 to 6. Fig. 1(a) illustrates the distribution of 50 N users across the other splits, and Fig.1(b) shows the costs of RRU, RCC, and FH, applying an ILP method. The objective function is to find a balance between the centralized and decentralized levels of the C-RAN. The centralized level is weighted by β and takes into account the RCC, PUE_C , and P_C. The decentralized level is weighted by α and takes into account the RRU, . PUE_U , and P_U . The traffic load on the fronthaul is taken into account by calibrating the weighting factor γ.

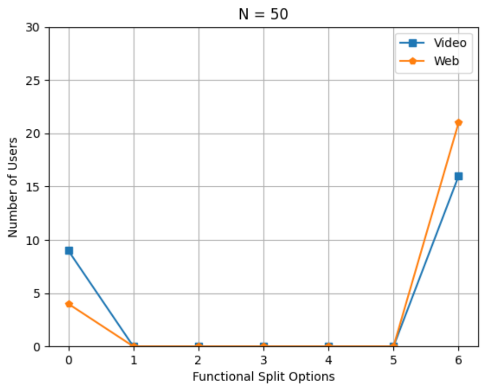


Fig. 1(a). User assigns in different split options

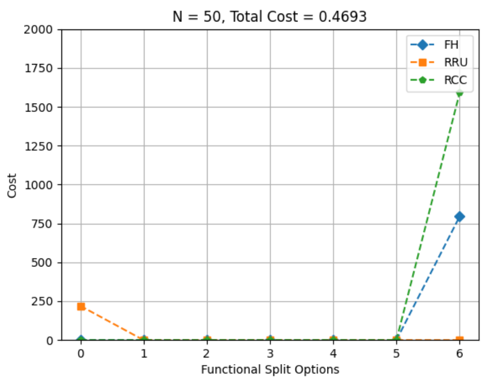


Fig. 1(b). Total cost and computational cost of RRU, RCC and FH

IV. Particle Swarm Optimization Approach

The optimization problem of the functional split, which was posed in the previous section III, is resolved in this section using the PSO algorithm. A population-based optimization technique, the PSO algorithm is motivated by the behavior of fish schools and bird flocks. Each particle in the population of particles used by the algorithm initially stands for one potential solution to the problem [13]. The three main characteristics of each particle in the PSO algorithm are position, velocity, and personal best (Pbest). While the velocity is the change vector that allows the particle to advance to the next position, the position represents the possible solution configuration. The particle's best solution configuration, as determined by its evaluation using a cost function, is stored in the Pbest memory function. The best solution configuration among all the best local solutions for particles is what we refer to as the global best (Gbest). It stands for the overall best answer any swarm particle has come up with. It compares the Pbest values of each particle in the swarm to arrive at this Gbest value and choose the configuration with the lowest cost.

The PSO algorithm enables particles to explore and optimize their solution configurations iteratively, seeking to converge towards the best feasible solution for the given problem, using these position, velocity, Pbest, and Gbest variables. After the algorithm has run through all steps, each particle iteratively works with the others to define its new velocity component [13]. (5) presents the procedure where the new velocity is created based on the old velocity of the previous iteration, Split P, Pbest, and Gbest. The coefficients C_1 and C_2 in (5) and inertia weight W are intended to enhance the process's randomness of evaluation. The random numbers r_1 and r_2 add a stochastic element to the velocity update equation, which helps prevent the swarm from converging too quickly to a local optimum [6, 13, 14]. The new position P is updated once the new velocity has been calculated using (6).

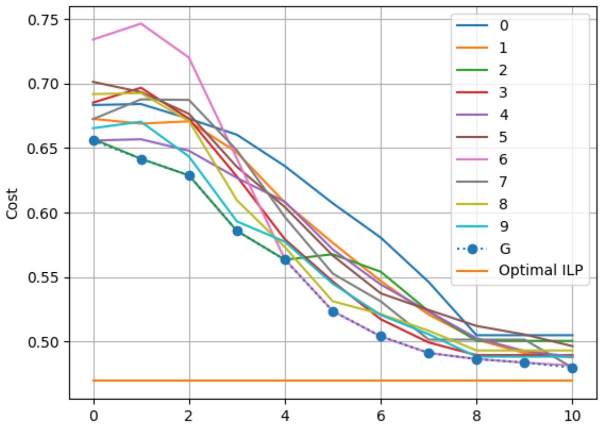
$$V_{new} = W.V_{old} + C_1 r_1 (P_{best} - Split_P) + C_2 r_2 (G_{best} - Split_P) \qquad (5)$$

$$Split_P = Split_P + V_{new} \qquad (6)$$

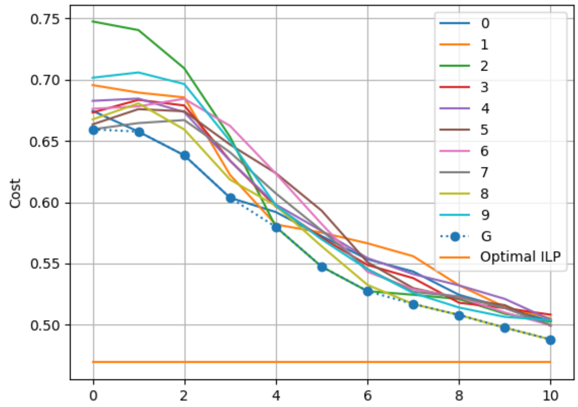
To initialize the random number generator with a predefined seed value for each execution in our implementation, we used the seed function. Controlling the seed value enables us to establish reproducibility in the generated random numbers, fostering consistent behavior across several program runs. The random number generator will reliably provide the same set of random numbers to this intentional initialization every time the application is run. PSO parameters are set to a population P of 10 particles and maximum iteration (*MaxIt*) is 10, C_1 and C_2 equal to 2. We assume the weight factor value for α = 0.8, β = 0.1 and γ = 0.1, respectively[14].

TABLE I. Simulation Parameters [16]

Parameters	Values
P (Particles)	10
N (Users)	50
MaxIt (Max Iteration)	10,15,20
K (Split Option)	7
α , β , γ	0.8,0.1,0.1
B	1228

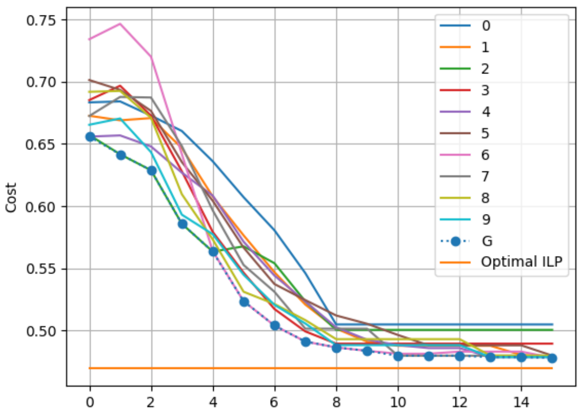


(a)Seed Value – 34342

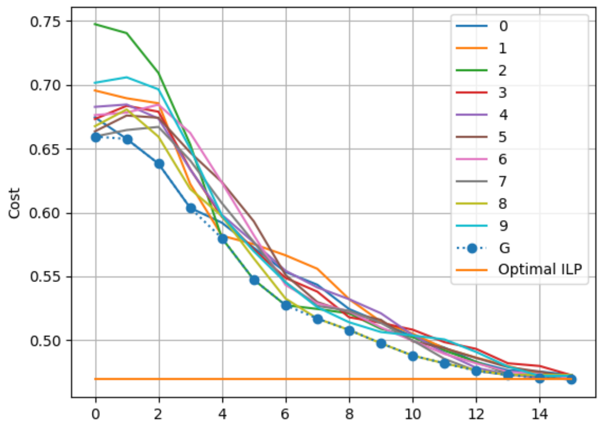


(b) Seed Value – 52454

Fig.2.Total Deployment Cost (MaxIt=10)

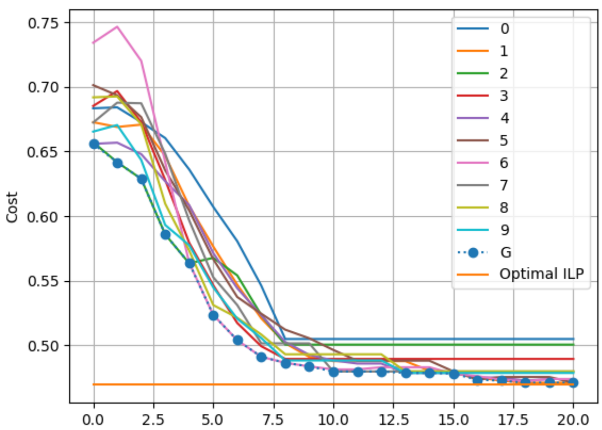


(a)Seed Value – 34342

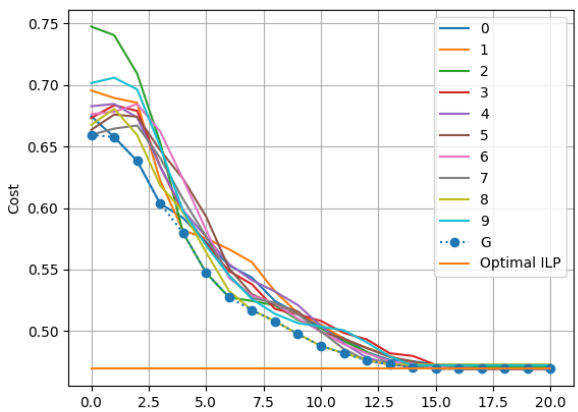


(b) Seed Value – 52454

Fig.3.Total Deployment Cost (MaxIt=15)



(a)Seed Value – 34342



(b) Seed Value - 52454

Fig.4.Total Deployment Cost (MaxIt=20)

We conducted several runs using various seed values during the optimization process to look at the impact of randomness on the results. In particular, we defined a maximum iteration number of 10, 15, 20 and carefully chosen two random seed values: 34342 and 52454.

TABLE II. Summarize the results of Deployment cost (P=10, N = 50)

Seed Value	Maximum Iterations		
	10	15	20
34342	0.4796	0.4779	0.4709
52454	0.4879	0.4693	0.4693

Table II offers a comprehensive summary of the deployment costs identified throughout the optimization process when N is set to 50. As we progressively increased the iterations to values of 10, 15 and 20, we observed a convergence in the deployment costs. These results were obtained through experimentation and analysis, allowing us to gain insights into the relationship between iterations and costs in our scenario. By visualizing this data, we can better understand the effect of varying iterations on the convergence of deployment costs.

CONCLUSION

Cloud Radio Access Network shows significant potential for cutting wireless network construction and maintenance expenses. The baseband processing tasks of a wireless network are centralized by C-RAN, which can drastically reduce the number of base stations needed and the equipment required at each base station[2]. This may result from significant cost reductions, enhanced capacity, and performance. The placement of RAN functions must be optimized to reduce computational costs and bandwidth utilization. Our method is based on Particle Swarm Optimization, a metaheuristic method that has effectively solved various optimization issues[13]. As we demonstrate by evaluating our approach, the overall cost of building a C-RAN network can significantly decrease. The outcomes of our research show that the PSO-based approach converges with the best options when comparing it to random search. Notably, the performance of PSO closely resembles the optimal result obtained by Integer Linear Programming as the number of iterations increases. This suggests that PSO can be a useful substitute for ILP for a range of optimization issues, particularly when the number of iterations is not a limiting constraint. We found that the deployment costs tend to stabilize and converge as the number of iterations expands through meticulous experimentation and observation. This result illustrates how crucial it is to optimize the iteration parameter in our deployment operations in order to get the desired cost results. We have shown the promise of the PSO-based technique as a promising method for optimizing the bandwidth as well as computational costs in C-RAN environments.

References

[1] k. Chen and R. Duan, “C-RAN: The Road Towards Green RAN,” Tech. Rep. V3.0, December 2013.

[2] “The Benefits of Cloud-RAN Architecture in Mobile Network Expansion.” White Paper,2014.

[3] Buschmann, Philippe, et al. "Task allocation in industrial edge networks with particle swarm optimization and deep reinforcement learning." *Proceedings of the 12th International Conference on the Internet of Things*. 2022.

[4] Selvaraj, Suganya, and Eunmi Choi. "Dynamic Sub-Swarm Approach of PSO Algorithms for Text Document Clustering." *Sensors* 22.24 (2022): 9653.

[5] A. Alabbasi and C. Cavdar, “Delay-aware green hybrid CRAN,” in Proc. 15th Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt), May 2017, pp.

[6] A. Checko, A. P. Avramova, M. S. Berger, and H. L. Christiansen, “Evaluating C-RAN fronthaul functional splits in terms of network level energy and cost savings,” J. Commun. Netw., vol. 18, no. 2, pp. 162–172, Apr. 2016.

[7] A. Pelekanou, M. Anastasopoulos, A. Tzanakaki, and D. Simeonidou, “Provisioning of 5G Services Employing Machine Learning Techniques,” in IEEE International Conference on Optical .

[8] A. Tzanakaki, et al., “5G infrastructures supporting end-user and operational services: The 5g-xhaul architectural perspective,” in IEEE Communications Workshops (ICC), Kuala Lumpur, Malaysia, May 2016.

[9] D. Harutyunyan and R. Riggio, “Flexible Functional Split in 5G Networks,” in IEEE International Conference on Network and Service Management (CNSM),2017.

[10] J. Liu, S. Zhou, J. Gong, Z. Niu, and S. Xu, “Graph-based framework for flexible baseband function splitting and placement in C-RAN,” in Proc. IEEE Int. Conf. Commun. (ICC), Jun. 2015, pp.

[11] S. Matoussi, I. Fajjari, S. Costanzo, N. Aitsaadi, and R. Langar,“A user centric virtual network function orchestration for agile 5G cloud-RAN,”in Proc. IEEE Int. Conf. Commun., May 2018.

[12] X. Wang, A. Alabbasi, and C. Cavdar, “Interplay of energy and bandwidth consumption in cran with optimal function split,” in IEEE ICC, Paris, France, May 2017.

[13] X. Wang, L. Wang, S. E. Elayoubi, A. Conte, B. Mukherjee, and C. Cavdar, “Centralize or distribute? A techno-economic study to design a low-cost cloud radio access network,” in Proc. IEEE Int. Conf. Commun. (ICC), May 2017, pp. 1–7.

[14] Y. Chen, J. Chen, G. Krishnamurthi, H. Yang, H. Wang and W. Zhao, "Deep Reinforcement Learning for RAN Optimization and Control," *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, 2021, pp. 1-6.

[15] J. Fang, W. Liu, L. Chen, S. Lauria, A. Miron, and X. Liu, “A Survey of Algorithms, Applications and Trends for Particle Swarm Optimization”, *IJNDI*, vol. 2, no. 1, pp. 24–50, Mar. 2023. doi: 10.53941/ijndi0201002.

[16] Salma Matoussi, Ilhem Fajjari, Salvatore Costanzo, Nadjib Aitsaadi, Rami Langar” 5G RAN: Functional Split Orchestration Optimization”, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 38, NO. 7, JULY 2020.