Indonesian Journal of Electrical Engineering and Computer Science
Vol. 40, No. 3, December 2025, pp. 1377~1390
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v40.i3.pp1377-1390 ❒ 1377

Dynamic resource allocation in cloud-radio access network using call detail record analysis

Régis Donald Hontinfinde[1,2], Ida Sèmèvo Tognisse[3], Marie Mélène Sèmèvo Tonou[1], Senan Ida Valeire Hontinfinde[1], Miton Abel Konnon[1]
[1]Laboratory of Science Engineering and Mathematics (LSIMA), National University of Science, Technology, Engineering and Mathematics (UNSTIM), Abomey, Republic of Benin
[2]Photonics and Digital Broadcasting Research Unit (UR-PHORAN), University of Abomey Calavi (UAC), Polytechnic School of Abomey Calavi (EPAC-UAC), Abomey Calavi, Republic of Benin
[3]Institute of Mathematics and Physical Sciences (IMSP), University of Abomey Calavi (UAC), Porto-Novo, Republic of Benin

Article Info

Article history:

Received Jan 17, 2025
Revised Oct 28, 2025
Accepted Nov 16, 2025

Keywords:

Big data analytic
Cellular network
C-RAN optimization
Deep learning
Telecommunications

ABSTRACT

We propose a solution based on call detail record (CDR) data analysis for cloud-radio access network (C-RAN) network optimization. First, we propose a data traffic prediction model in 3G and 4G networks using artificial intelligence (AI) models (long short-term memory (LSTM) and Bidirectional LTSM (BiLSTM)). Second, we propose a dynamic baseband units (BBU) resource allocation algorithm based on the obtained traffic prediction results to evaluate the rate of BBUs used as well as the average utilization rate of active BBUs in a C-RAN network. We used mean absolute error, root mean square error and mean absolute percentage error to evaluate the prediction model. The results obtained show that the best performance for estimating data traffic in 3G and 4G networks was obtained with the BiLSTM model, and is as follows: 1.143; 1.521; 2.47 percent for 4G, and for 3G, we have 0.2553; 0.3608 and 27.70 percent. Finally, evaluations with the predicted traffic dataset show that our framework provides up to 81% reduction in the number of BBUs used by the normal RAN. Moreover, active BBUs are exploited on average up to 88.34% of their capacity in a C-RAN compared to an average rate of 10.8% in a traditional RAN.

Corresponding Author:

Régis Donald Hontinfinde
Laboratory of Science Engineering and Mathematics (LSIMA), National University of Science Technology, Engineering and Mathematics (UNSTIM)
Abomey, Republic of Benin
Email: donald.hontinfinde@yahoo.com

1. INTRODUCTION

With the fourth industrial revolution, the use of smartphones and internet of things (IoT), the volume of network network traffic has increased exponentially in recent years [1]. To cope with the rapid growth of subscribers and increasing traffic demand, telecommunication network operators are deploying more and more base stations to extend their network coverage and adding more powerful processing units to increase their network capacity [2], [3]. The increasing traffic volumes therefore pose great challenges for cellular network operators to reduce operational costs and ensure quality of service (QoS). Therefore, designing cost-effective and quality-aware network architectures is now essential for network operation and research [4]-[7]. Cloud radio access network (C-RAN) is a promising solution to address the many challenges [5]. C-RAN is no longer a secret to the scientific community; it has been widely studied in the literature [7]-[17]. It offers many advantages. First, better network performance, providing higher throughput to users.

Second, resource sharing among different BBUs present in the pool, allowing to put some of these units in sleep mode when their workload is low. However, in order to fully exploit the power of the C-RAN architecture, one of the major challenges is to design appropriate mapping schemes between remote radio head (RRH) and BBUs, so as to maximize the utilization rate for the entire network [7]-[8]. However, most of these works have proposed dynamic resource allocation algorithms with computational complexity, which is the problem we address in our solution in this work, aiming at achieving the optimal cost and scheduling objectives, we propose a data-driven C-RAN optimization algorithm to address the above challenges. From the traffic prediction models developed using real network data, we were able to dynamically optimize the RRH-BBU mapping schemes for C-RAN. The method used can be seen in two phases. we propose a data traffic prediction model in 3G and 4G networks using artificial intelligence (AI) models (long short-term memory (LSTM) and Bidirectional LTSM (BiLSTM)). Second, we propose a dynamic baseband units (BBU) resource allocation algorithm based on the obtained traffic prediction results to evaluate the rate of BBUs used as well as the average utilization rate of active BBUs in a C-RAN.

2. METHOD

2.1. Related work

In the literature, several works have proposed prediction models for mobile data traffic [11], [12], [15], [18]-[25]; as well as dynamic resource allocation schemes between RRHs and BBHs [11]-[15]. With the digital revolution, a massive amount of big data on cellular networks, such as call detail records (CDRs), has been generated, providing researchers with new opportunities to understand mobile user dynamics and thus better plan resource allocation. The knowledge discovered from these big data can be used to guide the optimization of cellular networks. In the literature, several works have proposed prediction models for mobile data traffic [11], [12], [15], [18]-[25]; as well as dynamic resource allocation schemes between RRHs and BBHs [11]-[15]. But there is still a gap in the literature with regard to the provision of efficient algorithms for the optimal use of BBUs resources. In the remainder of this work, we will focus on this aspect using traffic data to simulate the results. The exponential growth of mobile data traffic, driven by the proliferation of smart devices and emerging applications, has necessitated innovative RAN architectures such as C-RAN and open-RAN (O-RAN). C-RAN centralizes baseband processing into a shared BBU pool, improving resource utilization and reducing operational costs [26]. O-RAN extends the C-RAN paradigm by introducing openness and intelligence through the RAN intelligent controller (RIC), enabling dynamic and software-based resource management [27]. Nevertheless, efficient resource allocation remains a central challenge, given fluctuating traffic patterns, user mobility, and stringent QoS requirements. This review summarizes recent advances in dynamic resource allocation strategies for C-RAN and O-RAN, emphasizing the transition from traditional optimization techniques to data-driven approaches.

In C-RAN, resource management is generally divided into computational resource management (CRM) and radio resource management (RRM) [28]. Traditional CRM methods focus on associating RRHs with BBUs to optimize computational efficiency. According to [26], clustering techniques may account for geographical proximity, traffic load, interference levels, QoS targets, or throughput maximization. Such strategies often employ bin packing or genetic algorithms, which perform adequately under static conditions but adapt poorly to rapidly changing network dynamics. RRM, by contrast, governs the allocation of radio resources such as spectrum and power. Common strategies include power control, joint optimization of multiple parameters (e.g., power and bandwidth), and sum-rate maximization [26]. While these methods form the basis of resource management in C-RAN, their reliance on predefined models limits their applicability to dynamic and heterogeneous environments.

The evolution toward O-RAN and cloud-native architectures has fostered the integration of AI and machine learning (ML) to overcome these limitations. By exploiting real-time network data, AI-driven solutions offer adaptive and context-aware resource management. For instance, [27] developed an ML-based xApp for dynamic physical resource block (PRB) allocation on the near-real-time RIC. Using a random forest classifier, the system selects among four allocation policies; Equal allocation, voice priority, mobile broadband (MBB) priority, and dedicated resource reservation; based on network conditions and QoS demands. Simulations in a 5G heterogeneous network (HetNet) achieved an 85% accuracy in policy selection, improving scheduling performance at the O-RAN distributed unit (O-DU). Similarly, [28] employed reinforcement learning (RL) for adaptive centralized unit (CU) and DU selection, using an actor–critic model to balance observability and latency under varying traffic and delay conditions. Results obtained with ns3-gym simulations showed reductions in latency and improvements in throughput and packet delivery compared to static deployments.

Further work by [29] extended deep RL to cloud-native wireless systems, addressing resource allocation in network slicing and multi-access edge computing (MEC). Two algorithms were proposed: Twin delayed deep deterministic policy gradient (TD3) for continuous bandwidth allocation and deep Q-network (DQN) for discrete task offloading. Experiments on a Free5gc-based testbed demonstrated superior adaptability and efficiency over static schemes, confirming the scalability of deep RL for complex O-RAN architectures.

Overall, conventional techniques remain relevant for their simplicity and theoretical rigor, yet they lack the flexibility required for dynamic network conditions. Emerging AI-based methods, as demonstrated by [27]-[29], leverage learning-based models to adapt resource allocation in real time, significantly enhancing latency, throughput, and resource utilization. Although these approaches incur additional computational cost and demand substantial training data, they mark a decisive shift from static optimization toward intelligent and adaptive network control in C-RAN and O-RAN environments. Future research should explore hybrid models combining ML and RL, address scalability in large-scale deployments, and tackle challenges like fronthaul constraints and user mobility, as noted by [26]. These advancements will be crucial for realizing the full potential of next-generation RANs.

2.2. Problem formulation

In this section, we present the formulation of our problem as a constrained optimization problem. We will highlight the objective function as well as the optimization constraints.

2.2.1. Decision variables

$x_{ij} \in \{0,1\}$: Binary indicator, equal to 1 if RRH i is connected to BBU j, and 0 otherwise; $y_j \in \{0,1\}$: Binary indicator, equal to 1 if BBU j is active, and 0 otherwise.

2.2.2. Parameters

$NRRH$: Total number of RRHs.; $NBBU$: Total number of BBUs; C_j: Maximum capacity of BBU j; T_i: Load (traffic) of RRH i.

2.2.3. Objective function

Minimize the number of BBUs used and maximize the average utilization rate of active BBUs,

$$\text{Minimize : } \alpha \sum_{j=1}^{N_{BBU}} y_j - \beta \frac{\sum_{j=1}^{N_{BBU}} \frac{\sum_{i=1}^{N_{RRH}} T_i \cdot x_{ij}}{C_j} \cdot y_j}{\sum_{j=1}^{N_{BBU}} y_j}$$

where,

- $\alpha > 0$ Weight for minimizing the number of BBUs used.
- $\beta > 0$ Weight for maximizing the average utilization rate of active BBUs.

2.2.4. Contraints

- An RRH can only be connected to one BBU,

$$\sum_{j=1}^{N_{BBU}} x_{ij} = 1, \forall i \in \{1,2, \dots, N_{RRH}\}$$

- A BBU can be connected to multiple RRHs, but its total load cannot exceed its capacity,

$$\sum_{i=1}^{N_{RRH}} T_i \cdot x_{ij} \leq C_j y_j, \forall j \in \{1,2, \dots, N_{BBU}\}$$

- Consistent activation of BBUs,

$$x_{ij} \leq y_j, \forall i \in \{1,2, \dots, N_{RRH}\}, \forall j \in \{1,2, \dots, N_{BBU}\}$$

- Variable domains,

$$x_{ij} \in \{0,1\}, y_j \in \{0,1\}$$

2.2.5. Compact formulation

$$\text{Minimize: } \alpha \sum_{j=1}^{N_{BBU}} y_j - \beta \frac{\sum_{j=1}^{N_{BBU}} \frac{\sum_{i=1}^{N_{RRH}} T_i \cdot x_{ij}}{C_j} \cdot y_j}{\sum_{j=1}^{N_{BBU}} y_j}$$

Subject to,

$$\sum_{j=1}^{N_{BBU}} x_{ij} = 1, \forall i$$

$$\sum_{i=1}^{N_{RRH}} T_i \cdot x_{ij} \leq C_j y_j, \forall j$$

$$x_{ij} \leq y_j, \forall i, j$$

$$x_{ij}, y_j \in \{0,1\}, \forall i, j$$

In the following section, we present the greedy algorithm that we proposed for solving our problem thus formulated.

2.3. Method

To provide an effective solution to this problem, we have adopted a multi-stage methodological approach: modeling and simulation using case studies and real data
We evaluated our algorithm on three levels and compared it to the current RAN network. The focus was on the number of BBUs used, the utilization rate of BBUs, and the utilization rate of active BBUs.

– Number of BBUs used

During the simulation period, we calculated the number of BBU cards used by our algorithm and compared it to the number used in the current RAN network.

– Utilization rate of BBUs

The BBU utilization rate (TB) is obtained by calculating the ratio between the number of BBUs used with our algorithm (TBCRAN) and the number used in the RAN network (TBRAN), multiplied by 100. The number of BBUs used and their utilization rate were evaluated at two levels: average level, the average value over the simulation period; Maximum level : the peak value during the entire simulation period. Lower values for these indicators imply better performance of the algorithm.

– Average utilization rate of active BBUs

At each hour, the average utilization rate of active BBUs was calculated by taking, for each BBU, the ratio of its load to its capacity, multiplied by 100, and then averaging this value across all active BBUs. The average utilization rate of active BBUs over the simulation period was obtained by averaging the hourly rates across the entire simulation period. The closer this indicator is to 100, the better the algorithm performs.

2.4. Traffic prediction

In the literature, LSTM and BiLSTM models have demonstrated promising performance, which influenced their selection for our study. In the following sections, we compare their results to determine which model achieves superior performance. For each RRH, and for both 3G and 4G, we implemented both models. To define the architecture of these models, we established value ranges for the hyperparameters. These ranges were kept identical for both models to ensure a fair comparison.

2.4.1. Prediction methodology

After acquiring and processing our data, our modeling followed two majors' steps,

– Training and validation phases :

In this first step, 70% of the data was used to train our different models. During training, 15% of the data served as a test set to validate the model and promote the convergence of the loss function.

– Test phase

During this phase, the performance of the best trained BiLSTM model is evaluated on the test data. For each sector, we used 15% of the data for testing purposes. Since the data is hourly, our mission is to determine the value of traffic for the next hour. We exploited the sequence of traffic values of the previous 24 hours to predict the value of traffic for the 25th hour. This was applied to all RRHs and for both chosen models.

2.4.2. Evaluation of models

We used three metrics to evaluate and compare our different models.

– RMSE: The square root of the mean square error, which is a relative error.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{1}$$

– MAE: indicates the average differences between the predictions and the actual values.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{2}$$

– MAPE: this is the percentage of the mean absolute error.

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{3}$$

3. RESULTS AND DISCUSSION

3.1. Proposed algorithm

3.1.1. Presentation of algorithm

We've proposed our algorithm in four steps. At each of these stages, a certain number of variables were utilized. The Table 1 presents all the variables used.

Table 1. Summary of the key variables used in the algorithm

Variables	Values	Types
pre_values	n previous values of traffic for all of RRHs	matrix
Traffic	values of predicted traffic for all of RRHs	vector
BBU_state	if BBU active, 1 else, 0	binary vector
commu_state	1 if RRH connected to BBU and 0 else	binary matrix
charge	charge calculated per BBU for $t+1$	vector

– Prediction step

This first step mainly involves predicting the traffic value for each RRH at time $t+1$. At this stage, our predictive model will be utilized by leveraging the traffic values from the previous 24 hours for each RRH. Thus, the model will take the previous traffic values as input and return the predictions for $t+1$. In our simulations for the subsequent stages of the algorithm, we used the actual traffic values already available. We employed 10% of our dataset to perform the simulation.

– Calculation of charge step

At time t, the variable commu_state provides information about the switching state between the BBUs and the RRHs. To calculate the load of each BBU at time $t+1$, we sum the traffic values of the RRHs associated with each of these BBUs.

– Overloaded BBU discharge step

After the load calculation, it often turns out that some BBUs are overloaded, meaning the traffic load they have to manage exceeds their actual capacity or the maximum load they can support. In this context, we offload some of the RRHs connected to these overloaded BBUs onto other BBUs following the procedure below.

First, for each overloaded BBU, we identify the RRH with the lowest traffic and attempt to offload it onto an active BBU that can fully handle its traffic without becoming overloaded. If this option is not feasible after exploration, the RRH is then simply offloaded onto an inactive BBU. Throughout this process, the variables BBU_state and commu_state are simultaneously updated.

– Adjustment step

The main objective of our algorithm is to reduce the utilization rate of BBUs within the BBU pool by aiming to deactivate as many BBUs as possible while ensuring maximum utilization of the active BBUs. In line with this objective, after the phase of offloading overloaded BBUs, we proceed to offload additional BBUs, particularly the least loaded ones, to optimize the utilization of the remaining active BBUs. This process is iterated until no further offloading is possible. Simultaneously, the variables BBU_state and commu_state are updated and returned at the end of the algorithm.

Here is our algorithm formulation.

Algorithm 1. Proposed algorithm

Input: Traffic at time $t, T_t = \{T_t^{RRH_i} \mid i = 1,2,\dots,N_{RRH}\}$
Capacity of BBUs C_{BBU_j};
Association state $Assoc_t(RRH_i, BBU_j)$;
Activation state $State_t^{BBU_j} \in \{0,1\}$
Output: Updated association $Assoc_{t+1}$ and activation state $State_{t+1}$.
Phase 1: Traffic Prediction
foreach RRH_i **do**
Predict $T_{t+1}^{RRH_i} = f(T_{t-23:t}^{RRH_i})$ // Using 24-hour historical data
end

Phase 2: Calculation of charge
foreach BBU_j **do**
Compute load: $L_{t+1}^{BBU_j} = \sum_{RRH_i \in Assoc_t(\cdot, BBU_j)} T_{t+1}^{RRH_i}$
end

Phase 3: Overloaded BBU Discharge
foreach BBU_j such that $L_{t+1}^{BBU_j} > C_{BBU_j}$ **do**
Identify RRH with lowest traffic:
$RRH_{min} = \arg min_{RRH_i \in Assoc_t(\cdot, BBU_j)} T_{t+1}^{RRH_i}$
Attempt reassignment to active BBUs:
foreach BBU_k such that $State_t^{BBU_k} = 1$ **do**
if $L_{t+1}^{BBU_k} + T_{t+1}^{RRH_{min}} \leq C_{BBU_k}$ **then**
$Assoc_{t+1}(RRH_{min}, BBU_k) \leftarrow 1$; break
end
end
if no active BBUs can accommodate RRH_{min} **then**
Assign to inactive BBU:
Activate $BBU_{inactive}$: $State_t^{BBU_{inactive}} \leftarrow 1$
$Assoc_{t+1}(RRH_{min}, BBU_{inactive}) \leftarrow 1$
end
end
Phase 4: Adjustment
Repeat until no further optimization is possible:
While rebalancing is feasible **do**
Identify least-loaded BBU:
$BBU_{min} = \arg min_{BBU_j \in \{Active\ BBUs\}} L_{t+1}^{BBU_j}$
Reassign its RRHs to other active BBUs, respecting capacity limits.
if all RRHs are reassigned **do**
Deactivate BBU_{min}:
$State_t^{BBU_{min}} \leftarrow 0$
end
end
Return $Assoc_{t+1}$, $State_{t+1}$

3.2. Prediction model

We followed the same modeling steps for all RRHs, both for 3G and 4G. Here, we will present the results obtained at each step.

3.2.1. Training and validation phases

To build robust models that fit our data well, we need to design an appropriate architecture for each of these models. Therefore, we defined value ranges and searched for combinations that optimize the chosen loss function. The value ranges used are the same for all RRHs, and they are presented Table 2.

Table 2. Hyperparameters and range of values for both models

Hyperparameters	Range
Number of layers	{3,4}
Number of neurons in layer	$20 \dots 150 \in \mathbb{N}_*$
Optimizer	Adam
Activation function	ReLu
Dropout	{0.1,0.2}
Number of epochs	{50}

3.3. Data

The data used in this study were collected from the radio sites of Celtiis Benin, specifically in the Abomey-Calavi district. This area was chosen for our study because it is one of the regions where network congestion issues are frequent. The area is covered by 25 sites, 24 of which are tri-sector, while the last site has an additional sector, making it a four-sector site. Celtiis Benin operates with three different technologies: 2G, 3G, and 4G. However, this study focuses exclusively on data traffic for 3G and 4G.

3.3.1. General data by site

For all the sites within the study area, there are 64 bandwidth units, also known as BBUs or universal BB processing units (UBBPs). Each site is equipped with a certain number of carriers, each managing a specific number of cells. The UBBPs cards come in three types of configurations: the first dedicated solely to 3G cells, the second exclusively to 4G cells, and the third capable of handling both 3G and 4G cells. The number of cells per carrier that each card can manage, depending on the configuration, varies based on the type of card. The following table provides a summary of these configurations by card category.

The Table 3 provides an overview of the number of cells supported by each type of UBBP card based on its configuration. Additionally, at the sector level for each site, three different types of RRHs (Ressource radio unit: RRUs) are used,

- RRU 5508 for low frequencies (700 MHz to 1,000 MHz).
- RRU 5502 for medium frequencies (1,700 MHz to 2,100 MHz).
- RRU 5301 for high frequencies (particularly 2,600 MHz in our case).

The distribution of these different RRH types across sectors is detailed in the following table.

The RRU 5508 and RRU 5502 manage both 3G and 4G cells, providing coverage for multiple technologies. In contrast, the RRU 5301 exclusively handles 4G data traffic and is available only in a limited number of sectors.

Table 3. Summary of the number of cells supported by each type of UBBP card according to configuration

UBBP	UMTS mode	LTE mode	UMTS-LTE FDD mode
UBBPe4	12	6	6-3
UBBPg1a	12	6	6-3
UBBPg2a	12	12	6-6 or 3-9
UBBPg2	12	12	6-6 or 3-9

3.3.2. Mobile traffic data

The data pertains to 3G and 4G mobile data traffic, recorded hourly over a four-month period, from April 1 to July 31, 2024. Traffic values were provided per cell for each site. To calculate the traffic per RRU, we associated each RRU with its corresponding cells. For the UMTS 2,100 MHz cells, the traffic was aggregated, as the RRU 5502 simultaneously manages all four cells. Thus, the traffic for this RRU corresponds to the combined total of the four cells. The mobile data traffic is measured in gigabytes (GB).

3.3.3. BBU capacities data

For this study, we opted to use the cards in LTE FDD mode, as 4G accounts for the largest share of mobile data traffic. The current capacities are provided in terms of the number of cells. Therefore, it is necessary to convert these capacities into the same unit as mobile traffic. The data transfer rate or speed was provided for three 4G cells. The transfer rate, which represents the amount of data transmitted per second, is expressed in megabits per second (Mbit/s).

Let V be the data rate, Q the amount of data, and T the time, we have,

$$V = \frac{Q}{T} \tag{4}$$

$$Q = V \times T \tag{5}$$

Since the data is hourly, we will aim to determine the amount of data that the cells can handle in one hour, $T = 1h$, so $T = 3600s$, which leads us to,

$$Q = 3600 \times V \tag{6}$$

The data rate being in Mbits, the amount of data will be in Mb (megabits), whereas the traffic data is in GB. 1 Gb = 1,024 Mb and 1 GB = 8 Gb, so this leads us to,

$$Q = \frac{3600 \times V}{1024 \times 8} \tag{7}$$

Let C be the capacity of a BBU in GB, n the number of cells that the BBU can support, and knowing that the data rate is given for three cells, we have,

$$C = \frac{n}{3} Q \tag{8}$$

where,

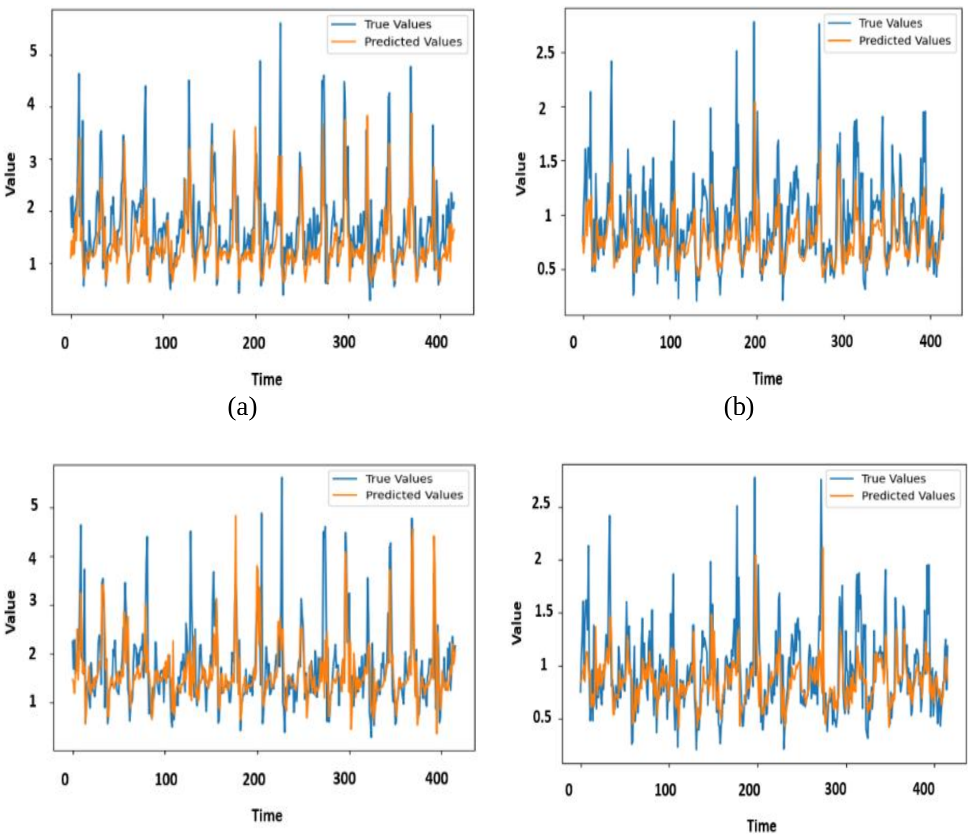
$$C = \frac{75 \times n \times V}{512} \tag{9}$$

with C en GB.

3.4. Test phase

3.4.1. 3G data prediction result

By comparing the actual values with the predicted values (Figure 1) for each of the different models applied to the two selected RRHs, we observe that the predicted values generally follow the trend of the actual values. Furthermore, the predictions made by the LSTM model (Figure 1(a) and Figure 1(b)), for the two selected RRHs appear to be lower than those of the BiLSTM model (Figure 1(c) and Figure 1(d)).



(a) (b)

(c) (d)

Figure 1. LSTM model vs BiLSTM model predictions at two RRH on 3G case; (a) LSTM prediction on RRH1, (b) LSTM prediction on RRH2, (c) BiLSTM prediction on RRH1, and (d) BiLSTM prediction on RRH2

Let us now focus on the prediction errors of the different models for the two selected RRHs. From Table 4, we can observe that for both selected RRHs, the BiLSTM model predominantly produces the lowest prediction errors (two out of three errors).

Table 4. Values of different metrics for 3G data - LSTM vs BiLSTM

	RRH1		RRH2	
Models	LSTM	BiLSTM	LSTM	BiLSTM
RMSE	0.706	0.668	0.380	0.360
MAE	0.487	0.443	0.273	0.255
MAPE	27.21	27.70	27.69	28.20

3.4.2. 4G data prediction result

As in the case of 3G, we observe that the predicted values from the different models for the two selected RRHs in the 4G case generally follow the trend of the actual values (Figure 2). Additionally, it is noteworthy that with the LSTM model (Figure 2(a) and Figure 2(b)), the predictions appear not to exceed a certain maximum value for the two selected RRHs. In contrast, the BiLSTM model does not exhibit this limitation and seems to have its predicted values (Figure 2(c) and Figure 2(d)) closer to the actual values compared to the LSTM model. Predominantly, the BiLSTM model also proves to be the one that yields the lowest errors in this case as well (three out of three errors) in Table 5.

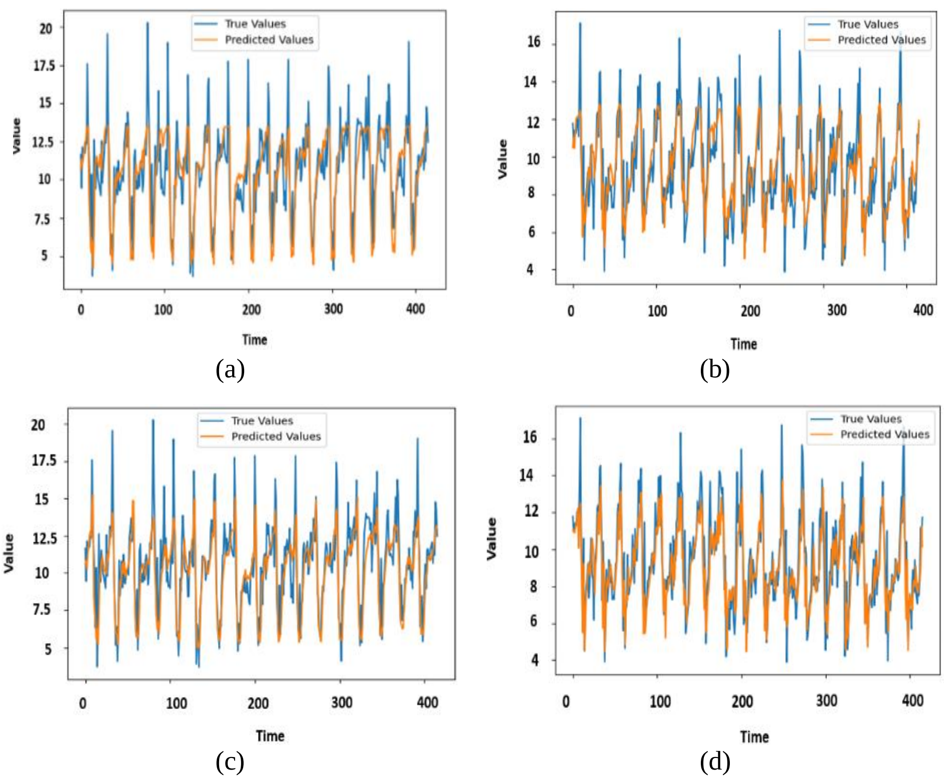


Figure 2. LSTM model vs BiLSTM model predictions at two RRH on 4G case; (a) LSTM prediction on RRH1, (b) LSTM prediction on RRH2, (c) BiLSTM prediction on RRH1, and (d) BiLSTM prediction on RRH2

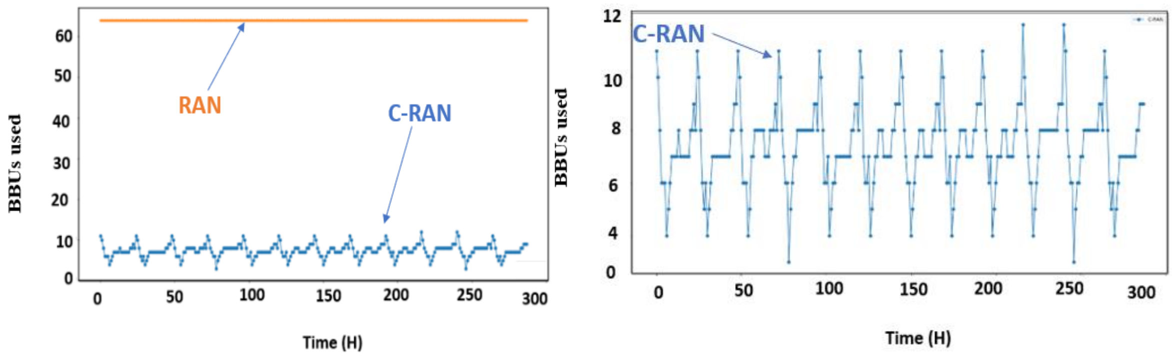
Table 5. Values of different metrics for 4G data - LSTM vs BiLSTM

	RRH1		RRH2	
Models	LSTM	BiLSTM	LSTM	BiLSTM
RMSE	1.804	1.633	1.536	1.521
MAE	1.375	1.209	1.200	1.143
MAPE	13.47	11.61	13.57	12.47

3.5. Dynamic resource allocation algorithm

To evaluate the performance of our algorithm on real data, we conducted simulations using 10% of our dataset. For the simulations, we assessed two scenarios: the current network architecture and our proposed algorithm. The simulation period spans 12 days.To calculate the traffic at each RRH, we summed the 3G and 4G data traffic at each moment for RRHs 5508 and 5502, as these two RRHs handle both 3G and 4G data traffic. For RRH 5301, we only considered 4G data traffic since it exclusively handles that. By comparing the number of BBU cards used in the current RAN network (Figure 3) with those of the C-RAN network with our algorithm, we observe that the number of cards in the current RAN network remains constant regardless of the time. In contrast, the number of cards in the C-RAN network with our algorithm varies (Figure 3(a)).

The number of cards used in the RAN is 64, whereas in the C-RAN network with our algorithm, it varies between 3 and 12 (Figure 3(a) and Figure 3(b)). We observed an average BBU utilization rate of 11.58% and a maximum utilization rate of 18.75%. This results in a reduction of up to 81% in the number of BBU cards used with our algorithm compared to the current RAN network. The statistics are presented in Table 6.



(a) (b)

Figure 3. Number of BBUs used - 12 days simulation (a) C-RAN Vs RAN and (b) C-RAN

Table 6. Values of different metrics for 4G data - LSTM vs BiLSTM

Level	RAN		C-RAN	
	Average	Maximum	Average	Maximum
Number of BBUs Used	64	64	7	12
Utilization Rate of BBUs	100	100	11.58	18.5

Let us now focus on the utilization rate of active BBUs with our algorithm and the current network. From Figure 4, we can observe that with our algorithm, active BBUs are utilized up to 88.34% of their capacity. Meanwhile, active BBUs in the RAN network are, on average, utilized at only 10.6% of their capacity. With our algorithm, we achieve a reduction of up to 81% in the number of BBUs used. Additionally, the BBUs that are used operate, on average, at 88.34% of their capacity.

The results are significantly better than many traffic prediction models available. By comparing the results obtained with those of by [26], which are based on theoretical resource management models in the C-RAN without addressing traffic prediction solutions, the method proposed in this work uses the BiLSTM model to achieve a MAPE of 2.47% for 4G, which is much better than many existing methods that do not incorporate real-time traffic prediction or have higher error rates. This forecasting quality allows the proposed algorithm to proactively allocate resources based on demand, which is a significant advance compared to the static approaches encountered in the study by [26], [30], use unsupervised learning techniques on CDRs to classify traffic, but their approach is at most concerned with network planning without being able to rationalise real-time resource allocation. The research work described here goes beyond simple classification, since predicted traffic models are then used for dynamic allocation of BBUs, enabling BBU utilisation to be reduced by up to 81%, with an average BBU utilisation rate of 88.34%, much higher than the 10.8% of a traditional RAN system. The proposed method shows a gain in efficiency in terms of resource utilisation, providing an answer to one of the limitations of the study by Zhou et al. By proposing an xApp for allocating optimal PRBs using a ML approach which, for example, displays a good 85% accuracy rate, by [27], do not integrate the dimensions of resource optimisation at the BBU level, which is conditioned by large-scale implementation considerations, not all of which are able to offer a complete resource management solution. By lightening the operation of the BBUs while achieving suitable utilisation rates for the active BBUs, one of the solutions we believe could, for example, generate significant savings while helping to ensure the efficiency of the network end-tool. The recommended method improves both latency and resource utilisation on the largest possible scale, based on real-time traffic forecasts and optimised BBU allocation.

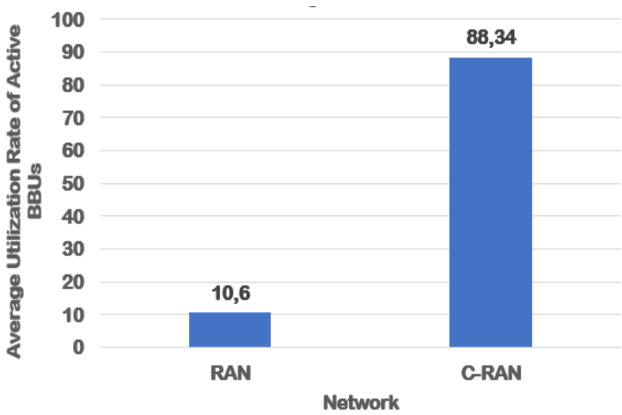


Figure 4. Average utilization rate of active BBUs - RAN vs C-RAN

4. CONCLUSION

Our study proposes a novel approach for optimizing cloud-RAN networks, using AI models to predict traffic and dynamically allocate BBU resources. This method significantly reduces the number of BBUs required (up to 81%) while maximizing their utilization, with an average utilization of 88.34% of active BBUs, compared to only 10.8% in a traditional network. These results confirm that data-driven traffic prediction combined with adaptive resource allocation can significantly enhance network efficiency, offering a scalable and low-cost optimization strategy. This study represents the starting point for a series of investigations. Future research may extend this work by leveraging the proposed traffic prediction framework based on LSTM and BiLSTM architectures to design more advanced resource management solutions. Integrating 5G, and eventually 6G, technologies, together with real-time and multi-source datasets, would further enhance the model’s applicability to next-generation networks. The proposed dynamic BBU allocation algorithm serves as a reference for intelligent resource management in shared or multi-operator environments. For experimental validation, future studies should focus on testing the model with real-world data collected from diverse scenarios, including urban and rural settings as well as peak traffic conditions. Additionally, evaluating the scalability and robustness of the model in large-scale simulated environments will be crucial for assessing its practical deployment potential. It is also essential to validate the system's robustness in the event of unforeseen events (sudden spikes, network failures) and to measure their impact on QoS. Finally, energy efficiency and economic viability studies should be conducted to strengthen the industrial value of the solutioneffective C-RAN adoption. In summary, by integrating AI into Cloud-RAN network optimization, this study paves the way for more efficient resource management, with significant gains in performance and reduced energy footprint.

ACKNOWLEDGMENTS

The authors thank the Beninese operator Cletiis for making all the data available for this study.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Régis Donald Hontinfinde	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Ida Sèmèvo Tognisse	✓	✓		✓		✓		✓	✓		✓	✓		
Marie Mélène Sèmèvo Tonou		✓			✓				✓					
Senan Ida Valeire Hontinfinde			✓	✓					✓	✓	✓	✓		✓
Miton Abel Konnon				✓	✓					✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data supporting the results of this study are available from the corresponding author, [RDH], upon reasonable request and with permission of the telecommunications operator Celtiis, which owns them.

REFERENCES

[1] Cisco Visual Networking Index, "Global mobile data traffic forecast update, 2016-2021," *Cisco Visual Networking Cisco*, pp. 2016–2021, 2017, [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf%0Ahttps://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-5208.

[2] P. Cerwall, A. Lundvall, P. Jonsson, S. Carson, and R. Moller, "Ericsson mobility report 2018," Sweden, 2018.

[3] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big data-driven optimization for mobile networks toward 5G," *IEEE Network*, vol. 30, no. 1, pp. 44–51, Jan. 2016, doi: 10.1109/MNET.2016.7389830.

[4] H. Suleiman, "A cost-aware framework for QoS-based and energy-efficient scheduling in cloud–fog computing," *Future Internet*, vol. 14, no. 11, p. 333, Nov. 2022, doi: 10.3390/fi14110333.

[5] China Mobile, "C-RAN: the road toward Green RAN," Beijing, China, 2011.

[6] A. Checko *et al.*, "Cloud RAN for mobile networks—a technology overview," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 405–426, 2015, doi: 10.1109/COMST.2014.2355255.

[7] H. Taleb, M. El Helou, K. Khawam, S. Lahoud, and S. Martin, "Centralized and distributed RRH clustering in cloud radio access networks," in *Proceedings - IEEE Symposium on Computers and Communications*, Jul. 2017, pp. 1091–1097, doi: 10.1109/ISCC.2017.8024671.

[8] K. Boulos, M. El Helou, K. Khawam, M. Ibrahim, S. Martin, and H. Sawaya, "RRH clustering in cloud radio access networks with re-association consideration," in *IEEE Wireless Communications and Networking Conference, WCNC*, Apr. 2018, vol. 2018-April, pp. 1–6, doi: 10.1109/WCNC.2018.8377287.

[9] N. Mharsi, "Cloud-radio access networks: design, optimization and algorithms," Université Paris Saclay (COmUE), 2019.

[10] I. S. Tognisse, A. D. Kora, and J. Degila, "Cloud-RAN and Coverage Gap in Rural Areas," in *10th IEEE International Conference on Communication, Networks and Satellite, Comnetsat 2021 - Proceedings*, Jul. 2021, pp. 329–336, doi: 10.1109/COMNETSAT53002.2021.9530816.

[11] L. Chen, T. M. T. Nguyen, D. Yang, M. Nogueira, C. Wang, and D. Zhang, "Data-Driven C-RAN optimization exploiting traffic and mobility dynamics of mobile users," *IEEE Transactions on Mobile Computing*, vol. 20, no. 5, pp. 1773–1788, May 2021, doi: 10.1109/TMC.2020.2971470.

[12] Y. Fu and X. Wang, "Traffic Prediction-Enabled Energy-Efficient Dynamic Computing Resource Allocation in CRAN based on deep learning," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 159–175, 2022, doi: 10.1109/OJCOMS.2022.3146886.

[13] B. J. R. Sahu, S. Dash, N. Saxena, and A. Roy, "Energy-efficient BBU allocation for green C-RAN," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1637–1640, Jul. 2017, doi: 10.1109/LCOMM.2017.2695545.

[14] W. C. Chien, C. F. Lai, and H. C. Chao, "Dynamic Resource Prediction and Allocation in C-RAN with edge artificial intelligence," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4306–4314, Jul. 2019, doi: 10.1109/TII.2019.2913169.

[15] J. L. Prepared and O. Falowo, "Machine learning based heuristic BBU-RRH switching scheme for C-RAN in 5G," Engineering and the Built Environment, 2019.

[16] N. Nikaein, "Processing radio access network functions in the cloud: critical issues and modeling," in *MCS 2015 - Proceedings of the 6th International Workshop on Mobile Cloud Computing and Services, co-located with MobiCom 2015*, Sep. 2015, pp. 36–43, doi: 10.1145/2802130.2802136.
[17] R. Tarek, "Virtualisation des fonctions d'un cloud radio access network(C-RAN)," Informatique mobile. Sorbonne Université, 2018.
[18] S. Jaffry and S. F. Hasan, "Cellular traffic prediction using recurrent neural networks," in *2020 IEEE 5th International Symposium on Telecommunication Technologies (ISTT)*, Nov. 2020, pp. 94–98, doi: 10.1109/ISTT50966.2020.9279373.
[19] V. Kurri, V. Raja, and P. Prakasam, "Cellular traffic prediction on blockchain-based mobile networks using LSTM model in 4G LTE network," *Peer-to-Peer Networking and Applications*, vol. 14, no. 3, pp. 1088–1105, May 2021, doi: 10.1007/s12083-021-01085-7.
[20] A. I. Abubakar, M. S. Mollel, and N. Ramzan, "FAMAC: a federated assisted modified actor-critic framework for secured energy saving in 5G and beyond networks," *arXiv preprint arXiv:2311.14509*, 2023.
[21] K. Sultan, H. Ali, H. Anwaar, K. P. Nkabiti, A. Ahamd, and Z. Zhang, "Understanding and partitioning mobile traffic using internet activity records data - a spatiotemporal approach," in *2019 28th Wireless and Optical Communications Conference (WOCC)*, May 2019, pp. 1–5, doi: 10.1109/WOCC.2019.8770653.
[22] M. E. Dodan, Q. T. Vien, and T. T. Nguyen, "Internet traffic prediction using recurrent neural networks," *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 9, no. 4, p. e1, Sep. 2022, doi: 10.4108/eetinis.v9i4.1415.
[23] T. Kuber, I. Seskar, and N. Mandayam, "Traffic prediction by augmenting cellular data with non-cellular attributes," in *IEEE Wireless Communications and Networking Conference, WCNC*, Mar. 2021, vol. 2021-March, pp. 1–6, doi: 10.1109/WCNC49053.2021.9417547.
[24] S. Sudhakaran, A. Venkatagiri, P. A. Taukari, A. Jeganathan, and P. Muthuchidambaranathan, "Metropolitan cellular traffic prediction using deep learning techniques," in *2020 IEEE International Conference on Communication, Networks and Satellite, Comnetsat 2020 - Proceedings*, Dec. 2020, pp. 106–110, doi: 10.1109/Comnetsat50391.2020.9328937.
[25] M. G. A. Pazou, R. D. Hontinfinde, R. T. Houessou, A. Agbodoyetin, J. L. C. Fannou, and C. Akowanou, "Traffic prediction of a mobile EDGE network using time series models: case of MTN Benin," in *2023 IEEE 10th International Conference on Communications and Networking, ComNet 2023 - Proceedings*, Nov. 2023, pp. 1–10, doi: 10.1109/ComNet60156.2023.10366612.
[26] R. T. Rodoshi, T. Kim, and W. Choi, "Resource management in cloud radio access network: conventional and new approaches," *Sensors (Switzerland)*, vol. 20, no. 9, p. 2708, May 2020, doi: 10.3390/s20092708.
[27] M. M. H. Qazzaz, Ł. Kułacz, A. Kliks, S. A. Zaidi, M. Dryjanski, and D. McLernon, "Machine learning-based xApp for dynamic resource allocation in O-RAN networks," in *2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, May 2024, pp. 492–497, doi: 10.1109/ICMLCN59089.2024.10625184.
[28] S. Mollahasani, M. Erol-Kantarci, and R. Wilson, "Dynamic CU-DU Selection for Resource Allocation in O-RAN using actor-critic learning," in *2021 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2021, pp. 1–6, doi: 10.1109/GLOBECOM46510.2021.9685837.
[29] L. Wang, J. Wu, Y. Gao, and J. Zhang, "Deep reinforcement learning based resource allocation for cloud native wireless network," *arXiv preprint arXiv:2305.06249v1*, 2023, [Online]. Available: http://arxiv.org/abs/2305.06249.
[30] J. Zhou, X. Liu, Y. Tao, and S. Yu, "QoS-aware power management with deep learning," in *2019 IFIP/IEEE Symposium on Integrated Network and Service Management, IM 2019*, 2019, pp. 289–294, doi: 10.1109/IM.2019.8748479.

BIOGRAPHIES OF AUTHORS

Régis Donald Hontinfinde is a lecturer at the University of Science, Technology, Engineering and Mathematics (UNSTIM) in the Republic of Benin. He graduated as a telecommunications design engineer from the Ecole supérieure Multinationale des Télécommunications (ESMT) in Dakar, Senegal in October 2013. In February 2018, he received a Ph.D. in telecommunications from the Polytechnic Faculty of the University of Mons in Belgium. His research interests include artificial intelligence, mobile communications networks, emerging technologies, optical fiber, applied informatics, algorithmic information theory. He can be contacted at email: donald.hontinfinde@yahoo.com.

Ida Sèmèvo Tognisse is computer and telecommunications engineer and Ph.D. in Information and Communication Technology from Institute of Mathematics and Physical Sciences at the University of Abomey-Calavi. She specialises in networks and telecommunications, and more specifically in digital broadcasting systems. For several years now, she has been involved in scientific research into digital inclusion in Africa. She can be contacted at email: idatognisse@gmail.com.

Marie Mélène Sèmèvo Tonou is an engineer in computer science, logistics and operational research who graduated from the National University of Sciences, Technologies, Engineering and Mathematics (UNSTIM) of Benin. She is passionate about modeling, artificial intelligence and operational research, grid computing, emerging technologies. She can be contacted at email: melenetonou@gmail.com.

Senan Ida Valeire Hontinfinde is a Senior Lecturer at the University of Sciences, Technology, Engineering and Mathematics (UNSTIM) in Benin and holds a Ph.D. in Theoretical Physics since 2016. She specializes in modeling and numerical simulation of complex physical systems, particularly in condensed matter physics and magnetism. Her research focuses on the modeling of dynamic systems, as well as on mathematical analysis and computational methods applied to interdisciplinary problems. She can be contacted at email: vhontinfinde26@gmail.com.

Miton Abel Konnon is an associate professor, lecturer at the University of Science, Technology, Engineering and Mathematics (UNSTIM) in the Republic of Benin. He holds a doctorate in computer science since 2011 and has a solid experience in the field of IT and Telecommunications, which he makes available to public and private universities in Benin. His research interests include artificial intelligence, mobile communications networks, IoT, optical fiber, information systems, applied informatics, and algorithmic information theory. He can be contacted at email: mdjidjohoako@yahoo.fr.