Optimizing the placement of cloud data center in virtualized environment

Yassir Al-Karawi, Raad S. Alhumaima, Khalid Hussein Khudair, Abdummunem Ahmed
Department of Communications, College of Engineering, University of Diyala, Diyala, Iraq

Article Info	ABSTRACT
	In cloud mobile networks, precise assessment for the position of the virtualization powered cloud center would improve the capacity limit, latency and energy efficiency (EEf). This paper utilized the Monte Carlo oriented particle swarm optimization (PSO) and genetic algorithm (GA) to first, obtain the optimal number of virtual machines (VMs) that maximize the EEf of the mobile cloud center, second, optimize the position of the mobile data center. To fulfil such examination, a power evaluation framework is proposed to shape the power utilization of a virtualized server while hosting an amount of VMs. In addition, the total power consumption of the network is examined, including data center and radio units (RUs). This evaluation is based on linear modelling of the network parameters, such as resource blocks, number of VMs, transmitted and received powers, and overhead power consumption. Finally, the EEf is constrained to many quality of service (QoS) metrics, including number of resource blocks, total latency and minimum user's data rate. *This is an open access article under the CC BY-SA license.*

Corresponding Author:

Yassir Al-Karawi
Department of Communications, College of Engineering, University of Diyala
Alqudus cross, Diyala, 23001, Iraq
Email: yassirameen22@gmail.com

1. INTRODUCTION

Aiming to provide multiple times higher spectral and energy efficiency (EEf) in the next generation systems, mobile administrators deployed small coverage cells individually or within a heterogeneous system. While this has expanded the network's capacity, it has likewise, prompted utilizing more power. This means causing releasing more harmful gases. To decrease this consumption, cloud mobile networks are proposed. In cloud mobile, also called cloud radio access networks, the base band units servers (BBs) are cloudified in a centralized area, called data center or base band unit pool BB pool. Thus, the radio unit (RU) is left exceedingly straightforward at the cell site, with just electrical to optical converter, amplifier, and antenna [1]–[5]. These BBs are in charge of processing the higher layers and large portion of the physical layer signals, including base band, radio frequency, media access control (MAC) layer signals. However, cloud mobile contradicts the conventional long term evolutions (LTE) framework, in the latter, the functions of the BB are handled in the eNodeB at the cell site itself.

Consequently, cloud mobile decreases the operational (OPEX) and capital (CAPEX) expenses via diminishing the maintenance cost and minimizing the sites' visits [6]. In addition, taking away the BBs from the evolved NodeB (eNodeBs) to the cloud has brought several advantages, such as advanced coordination, utilizing the available bandwidth, controlling the traffic variation and reducing the required cooling. This results in reducing the total power consumption.

In spite of this paradigm has promoted the EEf, however, escalating the quantity of the shall expand the power consumption [7]. Subsequently, offering new solution is an essential matter. So as to meet it, the networks has grasped the use of network function virtualization (NFV) [8]. NFV has granted the capability to benefit more of the available cloud resources, decreases the maintenance cost and activate the multi-tenant method of service. In which, each VM runs the BB functions and shares the assets of the host server with different VMs in a timely limited way. Running several VMs on the server demands a coordinator, called Hypervisor (HV), it gathers the data of all VMs regarding the number of users (UEs) each VM is capable to serve for specific time, in addition to their QoS requirements. Truth be told, both cloud networks and NFV speak to the key achievement advancements in the coming ages of networks generations [3].

To this level, several works have been proposed to make easy such adaptation. A low complexity, efficient and virtualization oriented resources scheduler has been proposed in [5] to maximize the throughput of the network users. Another scheduling process was proposed in [9] for transmitted power minimization with delay satisfaction for the users, similarly in [10]. However, in [11], both throughput and delay are optimized. Subsequently, the coexistence of the VMs inside the mobile server rises the server's RAM accessing, central processing unit functions complexity and hard drive usage. Consequently, increases the power overhead and latency of the host server. Note that both virtualization and cloudification are excellent candidates for the next generation networks. Yet, the consequences of these technologies are not yet fully tested. Hence, this paper evaluates the trade-offs of these technologies up on the power consumption, latency, processed resource blocks, data rate and energy efficiency. Furthermore, knowing the position of the virtualized cloud centers should perfectly help designing networks with low average delay, improved interference levels and more energy efficient. To fulfil these evaluations, the following contributions have been made:

Contributions: i) by utilizing genetic algorithm (GA), particle swarm optimization (PSO) and Hybrid GA-PSO, the position of the data center, together with the number of VMs are optimized; ii) we proposed to measure the traffic volume (number of RUs, channel gains, transmitted power) using the Monte Carlo, adapted inside the main algorithm to cover massive number of network traffic snaps while using poison point process (PPP) distribution for the RUs; and iii) the way VMs and resource blocks influence the power consumption of the servers is modelled. This modelling provides a logical assessment to the power consumption at the unit level.

Virtualization trade-off: virtualization technology is capable of reducing the increment in the power consumption [12]. Such power reduction is achieved by running the VMs on the host server. This is conceivable when the server is running off-line based applications where the latency constraint is relaxed. However, a virtualized server with 1 VM may take around multiple times more execution time to process the information compared to bare servers [13]. Such delay, when running on-line services, cannot be maintained so easily. Henceforth, optimizing the quantity of the installed VMs in one host server is essential. In addition, the delay of BB functions is direct proportional to the coming resource blocks. This implies when the scheduled resource blocks to each VM is increased, more delay is produced. This requires the number of allocated resource blocks for each VM must be optimized too. On the other side, the host server increases its power consumption as its components are now fully utilized. Thus, a virtualized server might consume about 40% extra power when compared to conventional partners [13]. Finally, since optimizing the delay and power consumption is a crucial to maintain an acceptable level of QoS, optimizing the position of the virtualized cloud center is further necessary due to following reasons: i) it ensures reduced distances to all the connected RUs, which results in less link delay; ii) less distance requires less transmitted power from the data center to the RUs. This in turn reduces the required transmitted power to the UEs and saves the energy; and iii) less delay means less path loss attenuation, this means maintaining accepted level of the transmitted signals to the UEs. This implies less consumption due to amplification process, which saves the energy.

Related research: in [14], an on-request VMs migration algorithm was proposed for disseminated server farms. These works aimed to diminish the carbon dioxide (CO2) footprint. Zhani et al. [15] have proposed live migration techniques among the servers to adjust the traffic fluctuation dynamically. The problem of VMs placement in the data center was also discussed in many researches [16], were the placement of VMs can be subjected to the amount of traffic to reduce the power consumption of the data center, improving the scalability and offering higher data rates. Furthermore, Wu et al. [17] considered the issue of VMs optimization over servers that are spread over distributed clouds using heuristic algorithms. To realize, the cost of power during the migration process of a VM may reach about (10 W and 32 W) in the destination and source servers, respectively [18]. Moreover, the placement of VMs in mobile data center was presented in [19], where the network traffic was evaluated using an EEf functions and then optimized using heuristic algorithms. However, the positioning of the data center itself was not examined. Alhumaima et al. [20] have optimized the number of active VMs under the restriction of link capacity limit, to reduce the cost of cloud mobile network. Yet, these works have further ignored the problem of positioning the data center and EEf evaluations. To follow, Alhumaima et al. [21] have discussed the issue of placing the BB pool. However, this case was without extensive consideration of UEs' resources allocation and virtualization. The non-virtualization case of the BB

pool placement in cloud environment was also discussed in [22]. Karneyenka *et al.* [23] proposed to centralized the BBs servers depending on the traffic pattern and mobility of the UEs. Likewise, the virtualized BBs placement problem was discussed in [24]. In addition, an optical-wise architecture has been proposed in [25] to support the virtualized service chains. For which, an approach based on optimal weighting for the non-coherent detection has been proposed in [26] for the purpose of minimizing the bit error rate and higher accuracy for the detection process. A 2000 km fiber transmission channel was proposed to transfer the generated information while improving the dispersion characteristic of the single mode fiber that can be used for the proposed method amongst the cloud center and RUs [27]. In addition, an energy-efficient approach based on utilization factor has been proposed for placing the virtual machines in the cloud centers using heuristic [28]. Finally, the problem of routing the traffic flow and virtual machines placement have been proposed in [29] to reduce the energy cost.

2. SYSTEM MODEL
2.1. Channel capacity and power allocations

The cloud network has been assumed to contain a number of RUs N. These RUs follow PPP type of deployment and coverage. Every RU n has a number of UEs Ue_n that are PPP placed with axes (x_{ue}, y_{ue}), each with small scale fading h. In addition, each UE ue is located at distance $d_{n,ue}$ from the RU n, where $d_{n,ue} = \sqrt{(x_n - x_{ue})^2 + (y_n - y_{ue})^2}$. The RUs are located at the coordinates (x_n, y_n), each RU is optical fiber connected and positioned with distance $d_{n,o}$ to the pool, where $d_{n,o} = \sqrt{(x_n - x_o)^2 + (y_n - y_o)^2}$, and the BB pool holds the location x_o and y_o to be optimised. In this manner, two strategies are proposed to pass the BB pool's power to the optical connected RUs. The first power distribution method is based on the distance, where RU will receive power depending on the distance $d_{n,o}$ to the pool. Meaning, the closer RU n is to the pool, the less power P_n^r it will get when contrasted with different RUs, i.e.,

$$P_n^r = ((P_p^t - L)\, d_{n,o})/(\sum_{n=1}^{N} d_{n,o}) \tag{1}$$

where P_n^r denotes the received power of the n-th RU, L signifies the losses of the fiber, and P_p^t is the transmitted power from the pool. To demonstrate, this model is based on the RU's distance, when the distance is less, it is directly proportional to the received power P_n^r, which lessens the latter on account that the total received power of all RUs is equal to the total power of the pool. In traditional network, the requirement for such power allocation (from the BB pool to RUs) is overlooked, in light of the fact that the BB processing unit is located inside the eNodeB. With cloud network, the BB unit is moved to the pool to generate the UEs' signals. The second method is based on both, the pool-RUs distance and channel gain of the UEs Ue_n. When the latter is high, and the RU is more far off to the pool, this allows an increase in the received power by the RU to compensate such attenuation, i.e.,

$$P_n^r = ((P_p^t - L)\, (h_{ue}^n\, d_{n,o}))/(\sum_{n=1}^{N} \sum_{ue=1}^{Ue_n} (h_{ue}^n\, d_{n,o})) \tag{2}$$

where ue denotes UE's index, and $h_{ue}^n = |h_{ue}|^2$ holds the attenuation of the UE ue located in RU n. It is worth mentioning that these methods can be converted to wireless power allocation if the loss L of optical fiber is removed from the above formula and replaced with wireless path loss. On the other side, the UE can be allocated an amount of power based on its distance $d_{n,ue}$, the received channel gain $h_{ue}^{n,k}$ and its path loss when compared to other UEs within the RU n, i.e.

$$P_{n,ue}^k = (P_n^r\, h_{ue}^{n,k}\, d_{n,ue}\, r_{ue}^n)/(\sum_{n=1}^{N} \sum_{ue=1}^{U_n} h_{ue}^{n,k}\, d_{n,ue}\, r_{ue}^n) \tag{3}$$

where $P_{n,ue}^k$ and $h_{ue}^{n,k}$ signify UE's ue power and gain from RU n that is served by VM k, $r_{ue}^n = (d_{n,ue})^{-\alpha}$ denotes path loss of the RU n to UE ue, while α represents the exponent of the path loss. Sequentially, the channel capacity is given as:

$$C = \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{ue=1}^{Ue_n} \sum_{b=1}^{B_k} B_o\, log\ _2(1 + P_{n,ue,b}^k\, \sigma_{n,ue,b}^k) \tag{4}$$

where B_k is the total resource blocks that is apportioned to the VM k at band-width B_o, $P_{n,ue,b}^k$ denotes power transmitted on resource block b. Furthermore, The SNIR is given as:

$$\sigma_{n,ue,b}^k = (P_{n,ue,b}^k \, h_{ue,b}^{n,k} \, r_{ue}^n)/(B_o N_o + I) \tag{5}$$

where $\sigma_{n,ue,b}^k$ denotes the signal to noise plus interference ratio (SINR) of the the resource block b that is allocated to UE ue, served by n-th RU that is connected to k-th VM, and N_o is the additive white Gaussian noise (AWGN) of the UE's channel. $I = \sum_{if \in \phi/n}^{If} h_{ue}^{if} \, r_{ue}^{if}$ is sum interference from different RUs (if) rather than serving RU n. In addition, $r_{ue}^{if} = (R_{ue}^{if})^{-\alpha}$ represents path loss of interferers RUs if of ue-th UE, and R_{ue}^{if} is distance set of interferers RUs if to ue-th UE, while h_{ue}^{if} is channel gain of if interferer to ue-th UE. It merits referencing that maximising the channel capacity for the UEs doesn't ensure that each UE's capacity is maximised separately. Henceforth, the capacity of every UE has been constrained, as displayed in (4). Further to channel capacity, the power consumption evaluation is the second part of the EEf problem, the following Sub-section presents modelling the cloud power consumption along with the expected delay.

2.2. Network energy consumption and latency

It was referenced in [30], [31] that the virtualized server's energy consumption is exponentially relative to the quantity of VMs. Nevertheless, there are four significant members inside the server required that participate to such assumption, the virtualized server power consumption is given as:

$$P_{server} = (P_{hd} + P_{nc} + P_{cu} + P_{rm}) \times e^{\xi K} \tag{6}$$

where K denotes the total number of VMs, P_{hd}, P_{nc}, P_{cu} and P_{rm} represent the primary power consumptions of the hard drive, network identification card, central processing unit and random access memory, respectively. The $e^{\xi K}$ part has been added to show the dynamic power consumption that is based on the traffic, which is increased because of VMs' presence, where ξ is a constant. Moreover, because each VM serves some UEs, it has been expected that the traffic based consumption is directly corresponding to the processed resource blocks [12]. This implies that more UEs being served, more processed resource blocks and higher dynamic consumption. Subsequently, the gross resource blocks processed by the server ($B = \sum_k^K b_k$) has been jointly combined with P_{server} to reproduce the power utilization ($P_{svr} = P_{server} + e^{\ell_k \times B}$). The amount b_k indicates the number of resource blocks handled by each VM k, and ℓ_k is a constant factor. It is worth mentioning that ℓ_k and the other constant factors are determined using try and error process, as it is the value that makes the power consumption of the server reaches its maximum consumption, starting from an initial value, i.e., when the server experiences no traffic.

Moreover, the power consumption of a virtualized server is exposed to the impacts of other power losses, i.e., cooling, DC and AC power conversion losses. The power consumption of these losses is scaled linearly with other units' power consumption and estimated as factors Λ_c, Λ_{dc}, Λ_{ac}, respectively [32]. Subsequently, total power consumption of the cloud virtualized network, denoted as P_{VC} can be modelled by combining the power consumption of cloud part, as in (7):

$$P_{VC} = \frac{(P_{svr} + P_{op,p})}{(1 - \Lambda_c)(1 - \Lambda_{dc})(1 - \Lambda_{ac})} + \frac{P_{amp} + P_{radio} + P_{op,r}}{(1 - \Lambda_{dc})(1 - \Lambda_{ac})} \tag{7}$$

where $P_{op,r}$ and $P_{op,p}$ symbolise the optical device's power consumption of the RU and the pool, respectively. Λ_{ac} and Λ_{dc} denote the RU's mains supply and DC loss factors, successively. At the end, P_{radio} is radio frequency component's power consumption, (P_n^t/η_{amp}) denotes the power amplifier's power consumption. To sum up, Figure 1 briefly shows how the system model parameters are related to each other and establish the EEf calculations.

Another important factor within this modelling is the time it takes the VMs to process the resources blocks. This time in traditional server is linearly increasing with the quantity of resource blocks and coding scheme utilized for resource blocks transmission. Concerning virtualization, a VM demands τ times extra delay to handle the packet when contrasted with the conventional partners because of the expanded computations among the hypervisor-server units and HV-VMs, where τ can reach 5 times [12]. Modelling this concept requires introducing a factor called coding scheme factor, denoted as M, to describe the linear relationship between the resource blocks and execution time in a bare BB server, denoted as G_{tr}, where $G_{tr} = G_{pri} + (M \times b_k)$, G_{pri} is the primary delay that is originated by other functions within the server, excluding the coding scheme. Moreover, the HV delay τ is added to G_{tr} to deliver the execution time of the virtualized server G^k when single VM can be found, that is, $G^k = G_{tr} + \tau$. Hence, the gross execution time for all the VMs, denoted as G is given as $G = \sum_k^K G^k$.
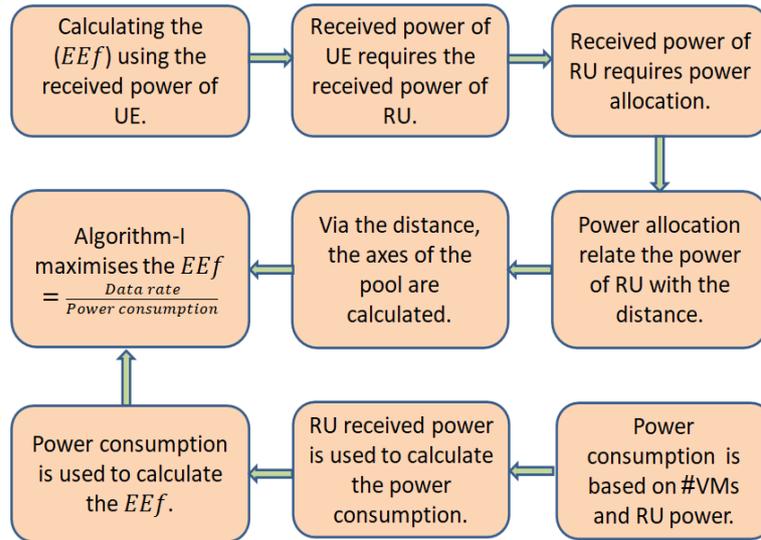
Figure 1. A flow diagram to show the EEf calculation through the system model

3. PROBLEM FORMULATION

The EEf (EEf) of virtual cloud network can be described as the received data rate in one watt. Such problem is formulated in (8) to (16):

$$\max \ \text{EEf} \ (K, x_o, y_o) = \frac{C}{P_{\text{VC}}} \tag{8}$$

$$\text{s.t} \quad C_{n,ue,b}^{k} \geq C_{th}, \ \forall \ ue, b \tag{9}$$

$$G_{ue} + G_n + G \leq G_{th} \tag{10}$$

$$G \leq G_{vms}^{thr} \tag{11}$$

$$b_k \leq 100, \ \forall \ k \tag{12}$$

$$\sum_{k}^{K} b_k \leq B \tag{13}$$

$$\sum_{ue}^{Ue_n} \sum_{b}^{b_k} P_{n,ue,b}^{k} \leq P_n^r, \ P_{n,ue,b}^{k} \geq 0, \ \forall \ ue, b \tag{14}$$

$$\sum_{n}^{N} P_n^r \leq P_p^t, \ \forall \ n \tag{15}$$

where the data rate can be given as

$$C_{n,ue,b}^{k} = B_o \log_2 (1 + P_{n,ue,b}^{k} \sigma_{n,ue,b}^{k}) \tag{16}$$

C_{th} represents the minimum data rate requirement. The constraint in (10) restricts the time delay due to processing the resource blocks and link delay, where $G_{ue} =* argmax(d_{n,ue}/c) \times 2$, represents the signal's round trip delay for the most far UE. Moreover, $G_n =* argmax(d_{n,o}/V_{ph}) \times 2$, is the maximum delay for the signal that travels from n-th RU to the pool, and G_{vms}^{thr} is the threshold time of all VMs. In addition, $V_{ph} = c/id$ is the speed of light (c) within the fiber, while id is its refractive index. Subsequently, G represents the delay of all VMs, which is equivalent to the sum of all VMs $\sum G^k$, where G^k is the delay of a single VM, where $G^k = G_{tr} + \tau$. The delay of processing the resource blocks can be evaluated by G_{tr} along with the VMs supervisor delay (τ). Hence, the third constraint combines the delay restrictions of processing the resource blocks and signals round trips.

By substituting the third constraint into the second, the latency for the network can be controlled and does not overcome G_{th}, where G_{th} is the total time threshold of the network. In constraint (11), any VM k does

not exceed the allowed number of resource blocks b_k. In addition, (12) manipulated the maximum processed resource blocks in one server. As processing capability of the server becomes larger and extra efficient by time, the total number of resource blocks has been assumed to be 700. Such number can then be shared by all participating VMs. The following constraint (13) expresses the limit of received power by the each UE to maintain the QoS. At the end, (14) ensures the received total power of all the RUs doesn't surpass the overall power transmitted by the pool P_p^t.

3.1. Heuristic algorithms

To solve the proposed problem, PSO, GA and HGAPSO are utilized to look through the space of the objective function to find the sub-optimal position of the virtualized BB pool x_o, y_o and the sub-optimal VMs K that increase the EEf given in (8). Such optimisation issue faces two main difficulties, i) the time limitations, it means for how long the solution is considered effective? and ii) the sub-optimality of the heuristic natured solution. Regarding first limit, considering a particular topographical zone where the pool existed, an immense measure of potential traffic has been considered. For every Monte Carlo cycle, new resource block assignments, channel conditions, RUs and UEs are set up by utilizing PPP. This iteration shall examine and cover all the possibility of traffic variations within the geographical area for a very large time period. Henceforth, running the process of optimization every time the UEs' traffic is varied becomes unnecessary.

The second limit in this problem is the results fall in to sub-optimal, but not optimal. This matter can also be ignored because of the nature of our problem. When the required amount of optimized VMs shall be an integer, the final solution is scaled down to overcome such behavior and to prevail server's overloading. The explanation of adjusting PPP at every iteration is to think about the practical cell coverage and genuine situations during the appropriation of UEs and RUs. The Poison method estimates the likelihood that a specific number of occasions happen inside a specific time-frame [31]. The heuristic algorithms primarily initializes their particles, with every particle speaking to a conceivable solution, such possibility at that point experiences the procedure of Figure 2, at which, the conceivable solution is exposed to the limitations.

In Figure 2 each particle i is evaluated by the given objective function EEf with previous best stored solution, denoted by $pbest$ and global best stored solution, denoted by $gbest$. In light of current positions of x_i, y_i, and K_i, their speed, denoted by v_i, their previous particle best solution $pbest_i$ and global particle best solution $gbest_i$, the particle will be updated until the maximum number of particles I is reached. In addition, the acceleration constants, denoted by c_1 and c_2, jointly with random vectors, denoted as r_1 and r_2, they all manage the stochastic impact on the speed of the particle [30]. Besides, the Monte Carlo procedure repeats J times at the i-th particle to produce $(I \times J)$ of assignment formations. When J is expanded, the quantity of traffic shots becomes nearer to endlessness and this will empower the activity of the results. In any case, this requires more execution time, which has been recently surpassed. All things considered; this time cost shall be increased when bigger geological area is deployed.

```
while (condition is not terminated) do Evaluate every initiated particle
for i, i = 1; 2; 3; I
do (Update best positions) if EEf (xi; yi; Ki) < f(pbsti),
pbsti = xi; yi; Ki

if f(pbsti) < f(gbsti), gbsti = pbsti,

end end

for j = 1: J,
Measure EEf (xij; yij; Kij)
end end

for i, i = 1; 2; 3; I (Next generation)

vi (t + 1) = wvi(t) + c1r1(pbsti (xi; yi; Ki)) + c2r2(gbst (xi; yi; Ki)),

(x i; yi; Ki) (t + 1) = (xi; yi; Ki) (t) + vi (t + 1) end

end while
```

Figure 2. Main PSO algorithm

4. RESULTS AND ANALYSIS

To correspond our results with experimental situations, the subsequent parameters were chosen from [30]–[32], as appeared in Table 1. These parameters established about 40\% power consumption increase inside each virtualized server because of server's over-using due to the presence of VMs [12]. Nevertheless, the proposed model is not limited to yield such amount of power. Rather, it is valid for any kind of server through tuning the model's inputs.

Figure 1 presents the EEf evaluation of cloud mobile network using distance-based power distribution from the BB pool to the RU. Due to the computational exertion that is required for PSO to arrive at its solution less than what is required by the GA to reach the same quality of solution, it was noticeable that PSO performance is better than GA. Moreover, in Figure 3 the EEf comparison has been shown using distance and channels gain from the BB pool to the RU.

Table 1. Model parameters

Components	Units	Values	Components	Units	Values
η_{amp}	-	0.36	ξ	-	0.007
id	-	1.5	ℓ	-	0.001
δ	-	0.9	P_{radio}	W	12.6
α	-	4	P_{nc}	W	2
b_n	-	90	P_{cu}	W	29.6
I	-	100	$P_{op,r}$	W	1
Λ_c	-	0.1	P_{hd}	W	10
Λ_{ac}	-	0.09	$P_{o,t}$	W	21
Λ_{dc}	-	0.075	C_{thr}	kbps	10
c_1	-	0.2	AWGN	dB/Hz	2620
c_2	-	1.2	G_{pri}	sec	80
v	-	0.005	G_{vms}^{thr}	ms	6
			L	dB/Km	0.51
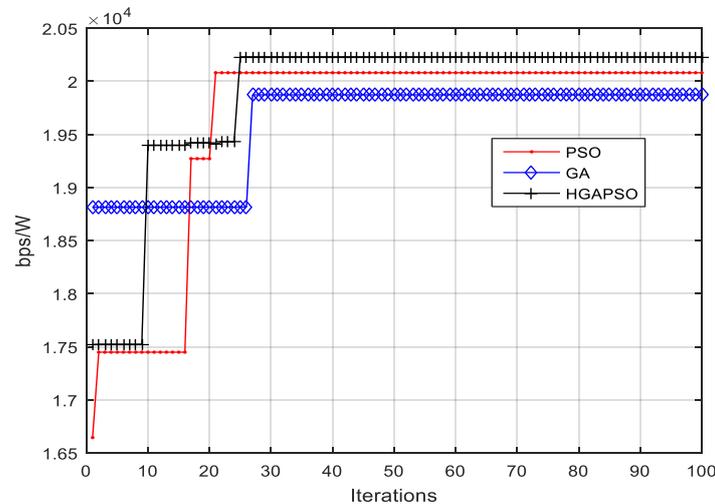
Figure 3. EEf evaluation of virtualization based cloud network using distance based allocation from the pool to the RU

It is worth mentioning that the first method, which is based on the distance shows better QoS than the second method, which is based on channel gain and distance, because of two main reasons:
− The channel gain (translated to attenuation) existence in the second allocation always degrades the EEf of the UEs as it reduces the received power by the RU, which in turn reduces the power received by the UE. Consequently, the SINR $\sigma_{n,ue,b}^{k}$ is reduced and caused reduction in the amount of data rate (channel capacity). As the EEf is directly proportional to the value of the data rate, the amount of EEf will be reduced too. This procedure is not allowed in the first method.
− The first method classifies the UEs to edge and center UEs within each RU. Since there is no channel gain assumptions, the EUs that are located at the center of the RU are always assumed to get good channel

conditions, while the UEs at the edge of the cell are assumed to always have poor channel conditions. Hence, extra power will be allocated to the edge UEs to compensate this poor channel. Subsequently, the SINR and channel capacity are increased, and this maximizes the EEf.

Figure 4 shows the EEf indicator when the network is in not under virtualization. Clearly, this case has delivered more power consumption, which is inversely proportional to the EEf, this degrades the latter as well as the QoS. To elaborate more on calculating the traditional power consumption, first the impact of VMs $e^{\xi K}$ is deducted from the server's power consumption P_{server}. If we assume $P_{bare} = (P_{hd} + P_{nc} + P_{cu} + P_{rm})$ is the server's power consumption without virtualization, it will be multiplied by the number of traditional servers, denoted as BU. Along with P_{RU}, i.e. $((BU \times P_{bare}) + P_{RU})$ they produce the total power consumption of Figure 5.

This amount of power has produced more power than in the virtualization case because traditionally, the bare servers are multiplied by the number of bare servers, while in the virtualization method, the server power consumption P_{server} only influenced by the number of VMs. Surely, the latter produces less power overhead than BU effect. However, the model results a value of the number of VMs between 7 and 8, this number then is down converted to 7. Additionally, Figure 5 combines all the proposed methods and techniques with respect to the EEf, including distance, and channel gain, using PSO, GA or HGAPSO, with virtualization (V) and without virtualization (NV).
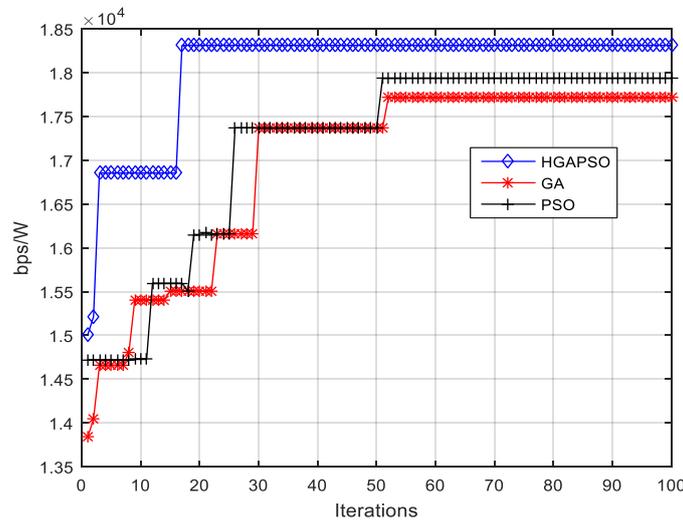
Figure 4. EEf of virtualization cloud mobile using channel and distance allocation from the pool to the RU
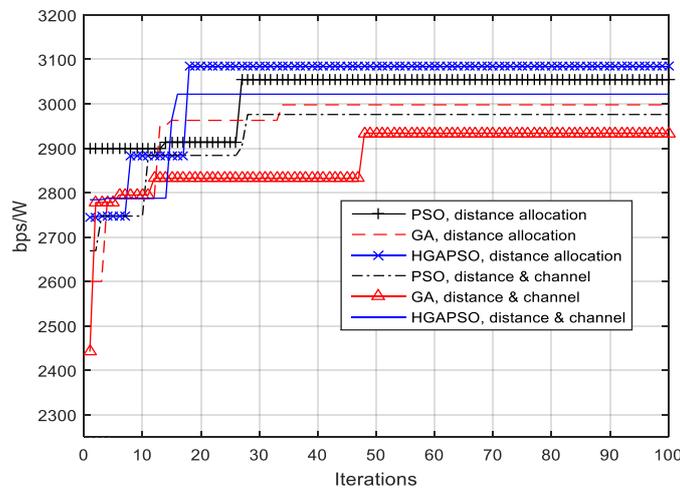
Figure 5. EEf performance without using virtualization method

The parameters for PSO are selected: the particles number I is 100, ($w = 0.91$) denotes the inertia weight, ($c_1 = 0.21$) denotes the cognitive factor, ($c_2 = 1.2$) is the social factor. In GA, generation number is also 100, size of population is equivalent to 100, and crossover probability is set to 0.81. The algorithm has been run 30 times to avoid the random behavior that is concurrent with the heuristic algorithm. Thereafter, the run that records higher is selected.

To further notice how the various parameters affect the optimization variables x_o, y_o and K, we have elaborated via the following steps, i) the EEf is based on the UE's received power, which depends on the power received from the RU that is received from the BB pool, ii) the received power of the RUs are evaluated using power allocation methods, iii) these methods also include the distances of the RUs-pool, which in turn, hold the axes of the RUs x_n, y_n and the optimisation variables of the BB pool position x_o, y_o, and iv) these axes then influence the received power of the RUs via the distances parameters, which affects the received power of the UEs and its data rate. As the data rate is affected, the EEf does too.

On the other side, the total power consumption is based on the received power of the RU that influences the power consumption of the power amplifier, as well as the number of VMs K. Hence, the latter will be optimised to a value so that the problem constraints will not be exceeded, in addition, the EEf and the QoS will not be degraded. If no constraints are found, K will be zero. This is because such value will minimise the total power consumption, which amplifies the EEf. However, two crucial limitations prevent such failure, i) the delay limit, which ties the VMs' latency to the value that is confined and ii) the total quantity of resource blocks is engaged with both the channel capacity and power consumption. The resource blocks plan to build the pace for channel capacity, at the same time, increases the power consumption because progressively executed resource blocks implies more consumed power.

5. CONCLUSION

The maximization problem of EEf in the virtualized based cloud mobile network is exhibited with regards to optimizing the position of the data center and the quantity of VMs in a single server. To empower such an assessment, the server's power consumption and transmitted power of the pool and RUs are modelled. These models quantify the consequence of expanding the processed resource blocks, the quantity of virtual machines and the transmitted power.

Moreover, the time limitation because of virtualization is displayed just as the execution time of the resource blocks in the servers. This plan is incorporated with the total latency of the cloud mobile network to take an interest in the optimization procedure so as the network's QoS will not be degraded. By adjusting Monte Carlo within the GA, PSO and Hybrid PSOGA in the area of interest, the need to run the procedure of optimization each time the traffic behavior changed is overridden. Finally, virtualization techniques and network planning are very useful in the long run when applied in cloud networks as they boost the EEf. But, they require extra caring with regards to fulfilling the UEs' QoS concerning the inherent delay. In the future, adapting this work to the newly coming software defined network architecture is motivating, where new assumptions that require new architecture suggestions, traffic flow design, resources allocation, complexity manipulation and costs estimation will be tested.

REFERENCES

[1] R. S. Alhumaima, M. Khan, and H. S. Al-Raweshidy, "Component and parameterised power model for cloud radio access network," *IET Communications*, vol. 10, no. 7, pp. 745–752, May 2016, doi: 10.1049/iet-com.2015.0752.

[2] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: a survey," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450–465, 2018, doi: 10.1109/JIOT.2017.2750180.

[3] M. Kalil, A. Moubayed, A. Shami, and A. Al-Dweik, "Efficient low-complexity scheduler for wireless resource virtualization," *IEEE Wireless Communications Letters*, vol. 5, no. 1, pp. 56–59, Feb. 2016, doi: 10.1109/LWC.2015.2495206.

[4] S. A. H. Mohsan and H. Amjad, "A comprehensive survey on hybrid wireless networks: practical considerations, challenges, applications and research directions," *Optical and Quantum Electronics*, vol. 53, no. 9, 2021, doi: 10.1007/s11082-021-03141-1.

[5] M. Kalil, A. Shami, and A. Al-Dweik, "Power-efficient QoS scheduler for LTE uplink," in *IEEE International Conference on Communications*, Jun. 2013, pp. 6200–6204, doi: 10.1109/ICC.2013.6655598.

[6] M. Khan, R. S. Alhumaima, and H. S. Al-Raweshidy, "Reducing energy consumption by dynamic resource allocation in C-RAN," in *2015 European Conference on Networks and Communications, EuCNC 2015*, Jun. 2015, pp. 169–174, doi: 10.1109/EuCNC.2015.7194062.

[7] R. S. Alhumaima and H. S. Al-Raweshidy, "Modelling the power consumption and trade-offs of virtualised cloud radio access networks," *IET Communications*, vol. 11, no. 7, pp. 1158–1164, Apr. 2017, doi: 10.1049/iet-com.2016.1396.

[8] Z. Chen *et al.*, "A survey on terahertz communications," *China Communications*, vol. 16, no. 2, pp. 1–35, 2019, doi: 10.12676/j.cc.2019.02.001.

[9] M. Kalil, A. Shami, A. Al-Dweik, and S. Muhaidat, "Low-complexity power-efficient schedulers for LTE uplink with delay-sensitive traffic," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4551–4564, Oct. 2015, doi: 10.1109/TVT.2014.2376038.

[10] M. Kalil, A. Shami, and A. Al-Dweik, "QoS-aware power-efficient scheduler for LTE uplink," *IEEE Transactions on Mobile*

Computing, vol. 14, no. 8, pp. 1672–1685, Aug. 2015, doi: 10.1109/TMC.2014.2363839.

[11] M. Kalil, A. Al-Dweik, M. F. Abu Sharkh, A. Shami, and A. Refaey, "A framework for joint wireless network virtualization and cloud radio access networks for next generation wireless networks," *IEEE Access*, vol. 5, pp. 20814–20827, 2017, doi: 10.1109/ACCESS.2017.2746666.

[12] R. Shea, H. Wang, and J. Liu, "Power consumption of virtual machines with network transactions: Measurement and improvements," in *Proceedings - IEEE INFOCOM*, Apr. 2014, pp. 1051–1059, doi: 10.1109/INFOCOM.2014.6848035.

[13] S. Bhaumik *et al.*, "CloudIQ: A framework for processing base stations in a data center," in *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*, 2012, pp. 125–136, doi: 10.1145/2348543.2348561.

[14] M. F. Zhani, Q. Zhang, G. Simona, and R. Boutaba, "VDC Planner: dynamic migration-aware virtual Data Center embedding for clouds," in *Proceedings of the 2013 IFIP/IEEE International Symposium on Integrated Network Management, IM 2013*, 2013, pp. 18–25.

[15] Q. Zhang, M. F. Zhani, M. Jabri, and R. Boutaba, "Venice: Reliable virtual data center embedding in clouds," in *Proceedings - IEEE INFOCOM*, Apr. 2014, pp. 289–297, doi: 10.1109/INFOCOM.2014.6847950.

[16] G. Wu, M. Tang, Y. C. Tian, and W. Li, "Energy-efficient virtual machine placement in data centers by genetic algorithm," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7665 LNCS, no. PART 3, Springer Berlin Heidelberg, 2012, pp. 315–323.

[17] D. Bhamare, A. Erbad, R. Jain, M. Zolanvari, and M. Samaka, "Efficient virtual network function placement strategies for Cloud Radio Access Networks," *Computer Communications*, vol. 127, pp. 50–60, Sep. 2018, doi: 10.1016/j.comcom.2018.05.004.

[18] Q. Huang, F. Gao, R. Wang, and Z. Qi, "Power consumption of virtual machine live migration in clouds," in *Proceedings - 2011 3rd International Conference on Communications and Mobile Computing, CMC 2011*, Apr. 2011, pp. 122–125, doi: 10.1109/CMC.2011.62.

[19] R. S. Alhumaima, R. K. Ahmed, and H. S. Al-Raweshidy, "Maximizing the energy efficiency of virtualized C-RAN via optimizing the number of virtual machines," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 4, pp. 992–1001, Dec. 2018, doi: 10.1109/TGCN.2018.2859407.

[20] J. Tang, W. P. Tay, T. Q. S. Quek, and B. Liang, "System cost minimization in cloud RAN with limited fronthaul capacity," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 3371–3384, May 2017, doi: 10.1109/TWC.2017.2682079.

[21] R. S. Alhumaima and H. S. Al-Raweshidy, "Optimising the BBU pool placement in cloud radio access networks based on power allocations," in *Proceedings of Computing Conference 2017*, Jul. 2018, vol. 2018-January, pp. 1312–1316, doi: 10.1109/SAI.2017.8252260.

[22] Y. Li *et al.*, "Joint optimization of BBU pool allocation and selection for C-RAN Networks," in *Optics InfoBase Conference Papers*, 2018, vol. Part F84-O, doi: 10.1364/OFC.2018.Th1B.5.

[23] U. Karneyenka, K. Mohta, and M. Moh, "Location and mobility aware resource management for 5G cloud radio access networks," in *Proceedings - 2017 International Conference on High Performance Computing and Simulation, HPCS 2017*, Jul. 2017, pp. 168–175, doi: 10.1109/HPCS.2017.35.

[24] C. Colman-Meixner, G. B. Figueiredo, M. Fiorani, M. Tornatore, and B. Mukherjee, "Resilient cloud network mapping with virtualized BBU placement for cloud-RAN," Nov. 2017, doi: 10.1109/ANTS.2016.7947790.

[25] V. Nikam, J. Gross, and A. Rostami, "VNF service chaining in optical data center networks," in *2017 IEEE Conference on Network Function Virtualization and Software Defined Networks, NFV-SDN 2017*, Nov. 2017, vol. 2017-January, pp. 1–7, doi: 10.1109/NFV-SDN.2017.8169845.

[26] W. Hu, Z. Wei, S. Popov, M. Leeson, M. Zhang, and T. Xu, "Non-Coherent detection for ultraviolet communications with inter-symbol interference," *Journal of Lightwave Technology*, vol. 38, no. 17, pp. 4699–4707, 2020, doi: 10.1109/JLT.2020.2993537.

[27] D. Semrau *et al.*, "Achievable information rates estimates in optically amplified transmission systems using nonlinearity compensation and probabilistic shaping," *Optics Letters*, vol. 42, no. 1, Dec. 2017, Art. no. 121, doi: 10.1364/ol.42.000121.

[28] E. Parvizi and M. H. Rezvani, "Utilization-aware energy-efficient virtual machine placement in cloud networks using NSGA-III meta-heuristic approach," *Cluster Computing*, vol. 23, no. 4, pp. 2945–2967, Feb. 2020, doi: 10.1007/s10586-020-03060-y.

[29] W. Fang, X. Liang, S. Li, L. Chiaraviglio, and N. Xiong, "VMPlanner: Optimizing virtual machine placement and traffic flow routing to reduce network power costs in cloud data centers," *Computer Networks*, vol. 57, no. 1, pp. 179–196, Jan. 2013, doi: 10.1016/j.comnet.2012.09.008.

[30] R. S. Alhumaima and H. S. Al-Raweshidy, "Evaluating the energy efficiency of software defined-based cloud radio access networks," *IET Communications*, vol. 10, no. 8, pp. 987–994, May 2016, doi: 10.1049/iet-com.2016.0046.

[31] Y. Liao, L. Song, Y. Li, and Y. Angela Zhang, "How much computing capability is enough to run a cloud radio access network?," *IEEE Communications Letters*, vol. 21, no. 1, pp. 104–107, Jan. 2017, doi: 10.1109/LCOMM.2016.2615612.

[32] G. Auer *et al.*, "How much energy is needed to run a wireless network?," *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40–49, Oct. 2011, doi: 10.1109/MWC.2011.6056691.

BIOGRAPHIES OF AUTHORS

Yassir Al-Karawi received a Bachelor degree in Electronics and Communications Engineering from the University of Technology, Iraq, in 2002. He worked as Engineer (2006-2010) at Diyala University. He received a M.Sc. in Communications Engineering from the University of Technology Malaysia (UTM) in 2012. He has published several journal and conference papers. Also, He is a Member of Staff with the Department of Communications, College of Engineering, University of Diyala, Iraq. His area of interest mainly focuses on radio frequency spectrum sharing optimizations, coexistence and compatibility studies between cellular and fixed services, including next-generation systems and networks. He is currently a PhD student at Brunel University London in Uk from 2020. He can be contacted at email: Yassirameen22@gmail.com.

Raad S. Alhumaima [ID] [g] [SC] [P] received his B.E. degree in Communications Engineering from the University of Diyala in 2003, his M.Sc. degree in Laser/ Electronics and Communication from the University of Baghdad in 2011, and his PhD in Communications and Electronics from Brunel University London-UK in 2017. He has Published several IET, IEEE and conference papers. His current research interests cover next generation networks, including SDN, C-RAN, NFV, Fog-RAN and Quantum mobile communication. He can be contacted at email: raad.1990.1990@gmail.com.

Khaled Hussein Khudair [ID] [g] [SC] [P] obtained a Bachelor's degree in Electrical Engineering from the College of Engineering at the University of Baghdad, Iraq in the year 1996. He obtained a master's degree in Electronic and Communications Engineering from the University of Baghdad, Iraq in the year 2000. Currently he works as a lecturer in Communications Engineering Department/College Engineering/Diyala University, Iraq. His field of research interests mainly focuses on communications, electronics and electronic power. He can be contacted at email: aljewary@gmail.com.

Abdel-Moneim Ahmed [ID] [g] [SC] [P] obtained a Bachelor's degree in Electronic Engineering from the University of Belgrade, Yugoslavia in 1984. He obtained a master's degree in Electronic Engineering from the University of Belgrade, Yugoslavia in 2000. Currently he works as a lecturer in the Department of Communication Engineering/College of Engineering /University Diyala, Iraq. His research interests are mainly in electronics and communications. He can be contacted at email: abdulmunem_alezy_eng@uodiyala.edu.iq.