IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 February 2024, accepted 6 March 2024, date of publication 11 March 2024, date of current version 15 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3375530

RESEARCH ARTICLE

Optimizing the Energy Efficiency Using Quantum Based Load Balancing in Open Radio Access Networks

YASSIR AL-KARAWI, (Member, IEEE), HAMED AL-RAWESHIDY, (Senior Member, IEEE), AND RAJAGOPAL NILAVALAN, (Senior Member, IEEE)
Department of Electronic and Electrical Engineering, Brunel University London, UB8 3PH London, U.K.
Corresponding author: Hamed Al-Raweshidy (Hamed.Al-Raweshidy@brunel.ac.uk)

This work was supported by the Department of Electronic and Electrical Engineering, Brunel University London, U.K.

ABSTRACT The open radio access network (ORAN) aims to improve network efficiency by reducing latency through the distributed unit (DU) and virtualization in the central unit (CU). However, the deployment of additional DU servers increases power usage and competition among virtual machines for user data processing. This study proposes a novel approach to address these challenges by incorporating load balancing strategies and quantum physics concepts, particularly entanglement theory in classical communication systems. The methodology enables the production of quantum information units from a single classical information unit, aiming to conserve energy. A simplified yet effective power consumption (PC) model for the ORAN architecture captures traffic fluctuations and provides a comprehensive characterization of power usage within a virtualized system. An optimization problem is formulated to select ORAN servers for quantum load balancing that are less energy-efficient, maximizing user benefits. The Lagrange multiplier method is used to deal with nonlinear objective functions. The presented problem is addressed through the use of two numerical methods, namely sequential quadratic programming (SQP) and the active-set approach. It is shown that the SQP model exhibits superior energy efficiency compared to the active-set model, with a difference of approximately 45%.

INDEX TERMS Open radio access network, quantum, load balancing, optimization.

I. INTRODUCTION

The open radio access network (ORAN) is a methodological approach that involves the dis-aggregation of mobile networks for their construction and operation. In conventional mobile networks, the radio access network (RAN) is characterised by its closed architecture, wherein the hardware and software elements are closely integrated and supplied by a sole vendor. ORAN, conversely, is founded upon open interfaces and standardised protocols that facilitate the interoperability and integration of components sourced from many suppliers [1], [2]. The concept of ORAN enables operators to strategically combine various components sourced from multiple vendors, hence facilitating the construction of a mobile network that is both adaptable and economically efficient. This phenomenon has the potential to result in heightened levels of competitiveness, reduced expenses, and accelerated rates of innovation. The utilisation of ORAN also facilitates the expeditious and effortless deployment of novel services and features by operators, as they possess the capability to carefully choose and seamlessly integrate the most suitable components that align with their particular requirements [3].

As the market for traditional mobile networks is primarily controlled by a limited number of significant manufacturers, ORAN has the potential to increased innovation and facilitate more favourable pricing structures for operators. It also possesses the capability to advocate novel services and features. One illustrative instance is enabling virtualized radio access network (VRAN), hence offering enhanced

The associate editor coordinating the review of this manuscript and approving it for publication was Wen Chen.

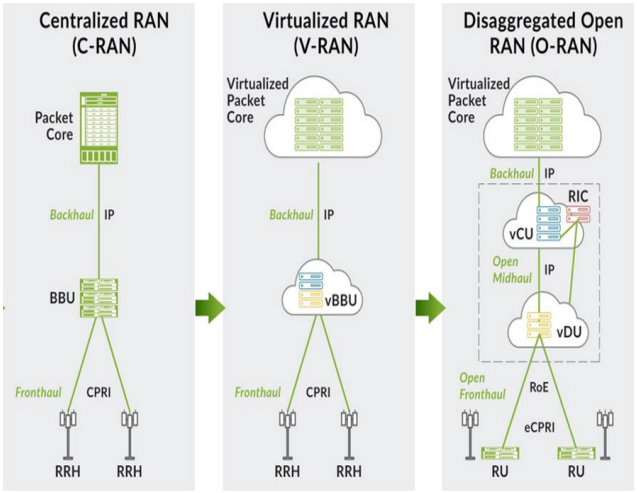

FIGURE 1. System architecture development of the ORAN.

utilisation of network resources and improved scalability [4]. The notion of ORAN is rather nascent, and efforts are now underway to establish standards and optimal methodologies for its implementation. The following information provides additional insights into ORAN:

1) The ORAN architecture is founded upon a software-defined methodology, wherein the RAN is disassembled into distinct functional elements that possess the capability to be autonomously deployed and administered. The connectivity of these components is facilitated through open interfaces and standardised protocols, hence facilitating interoperability across components sourced from various vendors [5].
2) The primary constituents of ORAN encompass the radio unit (RU), distributed unit (DU), and central unit (CU). The RU refers to the radio transmitter and receiver which establishes a connection with the user's device. The responsibility for processing and handling the radio transmissions lies with the DU. The responsibility for overseeing and organising the DU lies with the CU. Fig. 1 shows the system functions' upgrading from cloud to virtualized and then to ORAN [5].
3) ORAN is an incipient technology, however, there are collaborative network of manufacturers, operators, and industry associations that are collectively engaged in the development and implementation of ORAN networks. The growth of this ecosystem is seeing a rapid expansion, as an increasing number of vendors and operators are becoming part of the ecosystem on an annual basis [6].
4) The necessity for standardised interfaces and protocols, and the achievement of interoperability among diverse components sourced from various manufacturers is challenging. In order to tackle these problems, numerous operators and manufacturers are allocating resources towards ORAN. There also exist apprehensions regarding the security and dependability of ORAN networks due to their reliance on open interfaces and third-party components [7].

There are many research challenges and open areas regarding enabling ORAN, these can be summarised as follows:

1) The task of integrating and guaranteeing smooth interoperability among hardware and software components from several vendors and operators can present significant challenges.
2) The process of dis-aggregating the components of the RAN has the potential to result in heightened latency and performance concerns. The optimisation of communication and coordination between the DU and CU is of paramount importance [8], [9].
3) The increased number of interfaces and network units in ORAN architecture may render it susceptible to security breaches and malicious assaults.
4) The implementation of an ORAN infrastructure necessitates substantial initial financial commitments and entails intricate procedures for integration and testing.
5) The process of transitioning to ORAN necessitates meticulous preparation, rigorous testing, and effective collaboration in order to facilitate a seamless migration that does not impair the provision of current services.
6) The implementation of ORAN technology may necessitate regulatory modifications to effectively tackle challenges related to spectrum allocation and network ownership. The involvement of governments and regulators is crucial in the formulation of the policy framework for ORAN deployments [10].
7) On top of that, technically, the allocation and management of virtualized resources, such as virtual machines, latency, PC and bandwidth, can pose challenges in terms of efficiency. Efficient resource management algorithms and procedures are needed in order to maximise resource utilisation, yet without extra costs [11].

The ORAN architecture is specifically designed to enhance the proximity of some fundamental network operations to users by means of the DU unit. This will result in a decrease in the duration of connection and a decrease in latency. Furthermore, the virtualization process used within the CU and DU introduces an additional latency, as each virtual machine (VM) contends with others to process the data of its respective users [12]. Subsequently, the implementation of more DU servers results in an increase in the overall PC of the network. Therefore, it is essential to prioritise the reduction of PC in order to achieve optimal efficiency.

Classical solutions to network problems often compromise essential network metrics by focusing on one metric while neglecting others. For example, load balancing requires increasing power consumption, bandwidth requirements, and processing time. However, quantum mechanics offers benefits such as higher data rates, entanglement theory, and non-power-consuming optical equipment. Classical load balancing techniques include Round Robin, Weighted Round Robin, least connection method, least response time method,

dynamic weight adjustment, content-based routing, and machine learning-based load balancing. Each technique has its own limitations and advantages, and integrating multiple methods can enhance system performance and reliability while trading off among metrics. Thus, quantum methods are promising, advanced, futuristic, efficient, and low-cost methods compared to classical ones, enabling the solution of problems that cannot be solved using classical solutions.

This study introduces a tripartite resolution inside the framework of ORAN. The proposed solution involves the application of load balancing strategies by incorporating principles derived from quantum physics into classical communication networks, notably by utilising the theory of entanglement. The latter enables the generation of several units of quantum information with only one unit of classical information. The objective of this strategy is to preserve power within the cloud infrastructure. Furthermore, we have released a simplified yet effective PC model designed specifically for the ORAN architecture. This model is intended to precisely represent the fluctuations in server traffic volume. Moreover, it illustrates how PC might be characterised within the ORAN system. Furthermore, by employing this model, we established an energy efficiency optimisation problem. The primary goal of the establishment was to select the servers with the lowest energy efficiency for the load balancing procedure in the initial phase, for the advantage of the server users. The existence of nonlinearity in the objective function required the utilisation of the Lagrange multiplier method as a mathematical technique for solving the problem. The problem was addressed quantitatively using the 'fmincon' method in this work. Two distinct procedures were employed and subsequently evaluated for comparison. Two techniques, namely sequential quadratic programming (SQP) and the active-set approach, have been evaluated for solving the optimisation problem. Both of these strategies are specifically designed to handle objective functions that display traits of both nonlinearity and constraint behaviours.

This paper is structured as follows: Section II provides an overview of the quantum information and quantum entanglement and introduces some essential concepts and insights. The PC model is presented in Section III. The quantum PC model is provided in Section IV. Subsequently, in Section V, the optimisation problem pertaining to the selection of certain servers is outlined. The findings are presented in Section VII.

A. RELATED WORK

Given the inclusion of load balancing, PC modelling, quantum and optimisation problem in ORAN as the primary focal points of the proposed research, it is uncommon to encounter comparable studies. Consequently, the existing literature is segregated into several categories corresponding to these themes. The authors' objective was to investigate the most closely related research to the subject under consideration.

The work in [13] examines the adaptation cost associated with the utilisation of quantum technologies in classical communication. The problem of cluster head selection is addressed through the utilisation of the quantum approximation optimisation algorithm, with the objective of attaining an energy-efficient network. However, this work was not aimed for ORAN. The work of [14] provided an overview of the technological components involved in quantum computer systems, including quantum memory, quantum gate, quantum control, and quantum error correction. The investigation of the entropy of quantum channels is conducted in [15]. The authors of the [16] put forth a proposal for the implementation of a quantum repeater in order to mitigate network mistakes during the assessment of channel capacity. The study conducted by [17] involved the utilisation of a satellite to facilitate the interchange of entangled photons across a route spanning several hundred kilometres. In addition, the authors of [18] and [19] suggested models to utilise entanglement security in quantum internet networks. The utilisation of entanglement theory for the purpose of safeguarding network security through the implementation of quantum-based key distribution is explored in the study conducted in [20]. In the study conducted by Bartlett [21], it was demonstrated that both classical and quantum data transmission can be achieved without the need for partitioning the reference signal based on the photon's invisible sound transmission coefficient. The authors of the study conducted by [22] successfully transmitted high-dimensional quantum states over a distance of 2 kilometres using multi-core fibre.

Regarding ORAN, there still ongoing technical inquiries that seek to tackle its simultaneous obstacles and resources allocations. The existing body of study is constrained in its scope to encompass the examination of several aspects such as the difficulties encountered, progress made, structural framework, valuable perspectives, and proposed remedies. For instance, in [23], the ORAN architecture is presented to facilitate multi-vendor interoperability by utilising disaggregated, virtualized, and software-based components that are interconnected through open and standardised interfaces. The conventional RAN services are commonly characterised by their private and closed nature, leading to elevated expenses and a dearth of transparency for network operators. To effectively manage the limitations and minimise expenses, it is crucial to initiate a network redesign process that focuses on improving the efficiency of RAN installations and streamlining the operational aspects of RAN network services, as proposed in the study by [24]. The authors of [25] proposed an end-to-end network slicing method in multi-cell system. Nevertheless, the application of this technology in the RAN continues to provide significant challenges. In the study conducted by [26], it was shown that ORAN adopts cloudification and network function virtualization as methods for processing baseband functions. it showed that using heuristics enhances the economic efficiency and optimises the network resource in comparison to traditional greedy resource allocation algorithms. Furthermore, in [6],

the authors shed light on the existing constraints of the present ORAN standards and proposes potential technological solutions to address these restrictions, while the presented study of [27] focused on the development of a near real-time RAN intelligent controller service by the open networking foundation. The study also presents simulation findings pertaining to this service. Trends and opportunities are discussed in [4], advances in [28], and programmability of ORAN in [29].

In accordance to load balancing, the study conducted by Ye et al. [30] focused on the development of load balancing techniques within the RAN in the context of network slicing. In the study conducted by [31], an optimisation problem was presented with the objective of selecting the optimal split points for the ORAN. The primary aim is to achieve load balancing across CUs and midhaul links, while also taking into account the delay criteria. The formulation that arises from this problem is classified as NP-hard, and it is addressed using a heuristic algorithm. In a previous study [32], a load balancing technique was introduced with the aim of improving the overall sum-rate performance of the ORAN. Two sub-approaches were described that have the ability to function independently in a non-realtime RAN intelligent controller and a near-RT RIC, respectively. The findings demonstrated an improvement in the effective network sum-rate, along with enhanced load balancing among the radio units. A reinforcement algorithm was proposed in [33] for ORAN radio intelligent controller. load balancing in cloud radio access network can be found in [34], [35], and [36]. However, neither of these studies incorporated a quantum domain in their work, nor did they address the EE of ORAN.

In accordance to PC, the authors of [37] suggested a power model for assessing the performance of servers based on network demands. The power model indicated a non-linear relationship between the PC and the processing capacity as the load increased. The authors of the study proposed the utilisation of the ORAN architecture, as described in [38], to implement network functions in virtualized CUs and DUs by employing general-purpose central processing units in order to establish a processing pool. The allocation of this pool can thereafter be dispersed among different geographical networks, each with its own processing capacity, so influencing both network energy usage and performance. The study conducted by [39] provided a comprehensive analysis of different PC models, encompassing both virtualized and non-virtualized servers as well as data centres. The aforementioned models have been classified as intrusive, machine-learning, and software-based. However, the utilisation of intrusion-based models necessitates the deployment of intrusive tools and event counters, resulting in increased expenses and complexities associated with measuring PC performance. Software-based models necessitate the utilisation of supplementary software for their functioning, hence exhibiting a propensity for high PC and complexity. Machine learning algorithms are founded on heuristics, but this approach entails significant time and computational resource requirements.

According to [40], the study utilises a parameterized PC model to analyse the various components of the cloud radio access network, taking into account the number of virtual machines. The present study evaluates the performance characteristics and trade-offs associated with virtualizing a server. Before that, the Energy Aware Radio and Network Technologies (EARTH) PC assessment project introduces parameterized models for the state of the art base station in the works of [41] and [42]. However, these models were unable of accurately predicting the future power measure in hybrid networks, including virtualized and ORAN networks. In addition, the study conducted by [43] investigated the effects of virtualization on the performance of a single server's consumption when executing specific packages and apps. This study lacks presenting a mathematical model-based framework for evaluating the components or system level consumption, comparable to the approach taken in previous works in [44] and [45]..

One notable constraint observed in the majority of PC models discussed in the literature pertains to their specificity in terms of case selection, data type, and network configuration. Furthermore, the provided models are developed using exclusive intelligent software that is comprehensive and tailored to specific platforms, making it often inaccessible on an on-demand basis. Consequently, the utilisation of mathematical models that are both straightforward and capable of satisfying the needs of the general reader is the most effective approach for assessing the performance of personal computers. In general, there are some major falls within the existing works, first they are not suited for the ORAN design. In addition, some of the literature is based on the functions of each components, which leverage complexity. The other type are mathematical based models, but the model has no upper limits when the server is overloaded, such as in [12] and [46]

II. QUANTUM INFORMATION

A. QUANTUM BIT

The quantum bit, commonly referred to as the qubit, serves as the fundamental building block of quantum information. The system has the ability to exist in a superposition of two distinct basis states, which are represented as $|0\rangle$ and $|1\rangle$. Practically, a photon's polarisation of spin of electrons can encode the qubit. The quantum mechanical representation of a qubit's state can be stated as $|\psi\rangle$, and expressed as a linear combination of two basis states: $\alpha|\uparrow\rangle$ and $\beta|\downarrow\rangle$, as [47]:

$$|\psi\rangle = \alpha|\uparrow\rangle + \beta|\downarrow\rangle \tag{1}$$

where $|\uparrow\rangle$ and $|\downarrow\rangle$ denote the spin up and spin down of an electron; α and β be complex probability amplitudes that adhere to the normalisation requirement $|\alpha|^2 + |\beta|^2 = 1$. The chance of measuring the qubit in a particular state is obtained by squaring the absolute value of the probability amplitude.

B. QUANTUM ENTANGLEMENT

Quantum entanglement is a phenomenon where two or more quantum particles become correlated in such a way that their individual properties are no longer independent. The state of an entangled system cannot be expressed as a product of the states of its individual components, but must be described as a joint state using the tensor product of their Hilbert spaces. The state of an entangled system is described as [48]: $|\psi\rangle = \sum_{i,j} c_{i,j}|i\rangle_A|j\rangle_B$, where A and B represent the two particles that are entangled, $|i\rangle_A$ and $|j\rangle_B$ represent the possible states of each particle, and $c_{i,j}$ are complex probability amplitudes that satisfy the normalization condition $\sum_{i,j}|c_{i,j}|^2 = 1$.

For example, consider two spin-1/2 particles that are initially prepared in the state:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle_A|\downarrow\rangle_B - |\downarrow\rangle_A|\uparrow\rangle_B) \tag{2}$$

where $|\uparrow\rangle$ and $|\downarrow\rangle$ represent the two possible spin states of a particle along a given axis. This state is known as the singlet state and is maximally entangled. If we measure the spin of particle A and find it to be up, then we immediately know that particle B must be in the spin-down state. Similarly, if we measure the spin of particle B and find it to be up, we know that particle A must be in the spin-down state.

Another coding approach that is simpler for entangled photons is photon polarisation. It is noteworthy that the idea of entanglement presents two significant advantages in comparison to classical communication methods. Initially, it is observed that the quantity of created photons exceeds the quantity of classical bits employed in their generation. A single classical bit has the capability to stimulate a laser source, which in turn can provide the necessary energy to activate a nonlinear crystal. This activation process has the potential to produce a maximum of 12 entangled photons, out of one classical photon [49]. The second phenomenon pertains to the interconnections of these photons through a hidden channel. This implies that the transmission of information can occur seamlessly between the recipients of these photons once the state of one of the photons is modified. This feature provides a cost-free signalling procedure within the next generation of networks.

Quantum computing can offer several advantages over classical computing paradigms, particularly in the context of ORAN. It can perform computations on a massive scale in parallel, accelerating complex optimization tasks like load balancing [50], [51]. Quantum algorithms, such as quantum annealing and quantum approximate optimization algorithms, can also provide exponential speedups for optimization problems, improving efficiency in network performance [52]. Additionally, quantum computing offers advancements in cryptography techniques, such as quantum key distribution, which can bolster the security of communication networks [53]. Furthermore, quantum-inspired algorithms running on classical hardware can provide performance benefits, leveraging quantum principles to solve optimization and machine learning problems, potentially enhancing ORAN functionality [54].

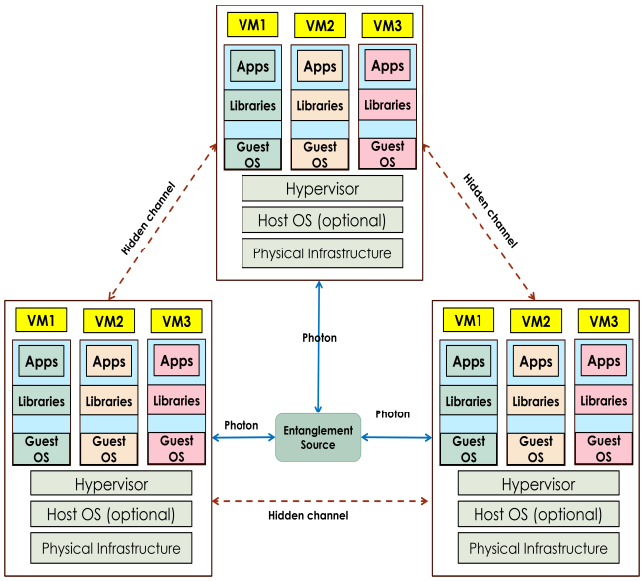

FIGURE 2. Proposed quantum based load balancing in ORAN.

The suggested method involves leveraging entanglement to facilitate the exchange of load balancing signalling information amongst the servers using photons as shown in Fig 2. Traditionally, the act of signalling between servers has been associated with significant costs in terms of PC and time [55]. Therefore, the suggested quantum approach provides a progressive means of information sharing across servers. In the proposed method, every server is assigned a photon, which is subsequently transformed into classical information. The state of the photon is then modified in order to disseminate this information simultaneously to the servers of interest. Traditionally, this necessitates the transmission of signalling information from each server to every other server individually to track the processing availability, resulting in increased PC and network latency. In order to achieve load balancing across the CU servers, we have measured the EE of the server's users as our criteria to decide the load balancing servers. Hence, it is essential to consider two primary metrics in this process: data rate and PC.

III. PC MODEL

It was assumed the PC of the network consists of cloud infrastructure that contains CU servers, DUs, RUs. In ORAN, there are many VMs that are responsible for enlarging the PC of the server up from its initial value to its maximum. Not only that, each VM is responsible for processing many resource blocks (RB), translated to bits, at each time slot. Moreover, the initial PC value of the devices is different amongst the servers. This usually based on the network vendor, manufacturer and device's characteristic. On top of that, it is not perfectly known how much power the software itself consumes, driving the idle power of the hardware [56]. Hence, offering detailed power model is a prerequisite to include the aforementioned parameters.

The modulation type, denoted as x governs the number of bits L_x in each resource element, for example, in QPSK, the number of bits is $L_x = 2$. In 16-QAM, $L_x = 4$, following the constellation of the modulation type, where $mod_x = 2^L$, is the number of bits in the constellation diagram. This number is multiplied to number of sub-carriers in the RB N_{sc}^{rb}, and number of symbols in the RB N_{sym}^{rb} to measure the number of bits in each RB, as follows [57]:

$$L_{rb}^{x} = L_x \times N_{sc}^{rb} \times N_{sym}^{rb} \tag{3}$$

To obtain the number of bits in all the RBs, the L_{rb}^{x} is multiplied by the number of RBs ($N_r b$), as follows:

$$L_{RB}^{x} = L_{rb}^{x} \times N_{rb} \tag{4}$$

On the other side, the power given in the RB is given in Eq. (5),which assumes that there is an equal allocation of power among all RBs, which guarantees a uniform distribution of total power resources without any preferential allocation. This process streamlines the model and facilitates mathematical computations, offering a fundamental scenario. Nevertheless, practical power allocation algorithms may differ, and real-world wireless communication systems frequently include more advanced control mechanisms to optimise performance. The assumption of equal power allocation facilitates comprehension of the model's behaviour.

$$P_{rb} = \frac{P_{NB}}{N_{rb}} \tag{5}$$

where P_{NB} and N_{rb} denote the power transmitted of the NodeB and the number of RBs, respectively. Note this is the power consumed in one time slot T_s. If we assume the total time of transmission is T, where $T = T_s \times N_{rb}$ where the transmission is discontinuous. In case of continuous transmission, $T = \int_{t_o}^{t_l} dt$, where t_o and t_l are the beginning and ending time of the transmission. Hence, the time averaged power allocated to the RB P_{rb}^{T} over a period of time T, is given by:

$$P_{rb}^{T} = P_{rb} \times T \tag{6}$$

which is assumed the total consumption of the VMs when processing a number of RBs. For one time slot, the VMs' PC is given by:

$$P_{vm} = \gamma P_{rb} \times N_{rb} \tag{7}$$

where γ denotes the share of the VM software of the total PC of the server. For example, if the hardware consumes 70%, then the software is 30%, where P_{vm} impacts. For a group of VMs, the PC can be modelled as:

$$P_{VM} = P_{vm} \times N \tag{8}$$

where N is the total number of VMs.

A. CU PC

Modelling the PC of the CU can mainly based on the existed VMs. However, the change in the PC of the CU varies according to the change in the number of VMs (N) and change of the time t at which these VMs are existed. It is worth noting that at any time instance, there are some VMs that are installed and increased the PC P_c within the server, where P_c denotes the consumed power at a certain time due to adding the VMs. In contrast, there are some of the VMs that terminating processing the UE's data and spared some power P_r to the total CU server PC P_{CU}. Whereas P_r, denotes the spared/gained power due to VMs' terminations. Meaning, the more spared power, the server can adapt more VMs and exploit the total consumption $P_C U$. Hence, the P_r is a function of $P_C U$, i.e., $P_r \propto P_{CU}$ or $P_r = \delta P_{CU}$, hence:

$$\frac{\partial P_{CU}}{\partial t} = P_c - P_r \tag{9}$$

or

$$\frac{\partial P_{CU}}{\partial t} = P_c - \delta P_{CU} \tag{10}$$

where δ is a constant, to follow:

$$\frac{\partial P_{CU}}{P_c - \delta P_{CU}} = \partial t \tag{11}$$

Integrating both sides reveals:

$$\frac{\ln(P_c - \delta P_{CU})}{-\delta} = t + c \tag{12}$$

where c is the constant of integration. The above expression can be written as:

$$P_c - \delta P_{CU} = e^{-\delta t} e^{-\delta c} = C e^{-\delta t} \tag{13}$$

If the initial condition is applied, the P_{CU} and t are replaced by P_{CU}^{o} and t_o, respectively. Hence:

$$C = (P_c - \delta P_{CU}^{o}) e^{\delta t_o} \tag{14}$$

Substitute Eq. 14 in 13, yields:

$$P_{CU}(t) = \frac{P_c}{\delta} + (P_{CU}^{o} - \frac{P_c}{\delta}) e^{-\delta(t - t_o)} \tag{15}$$

it is worth mentioning that the initial consumption is represented by P_{CU}^{o}. In addition, the model produces P_c/δ, which is the maximum value of consumption. To model the effect of VMs up on the CU consumption, replacing the time with number of VMs produces the same effect upon the P_{CU}^{o}, as follows:

$$P_{CU}(N) = \frac{P_c}{\delta} + (P_{CU}^{o} - \frac{P_c}{\delta}) e^{-\delta(N - N_o)} \tag{16}$$

where N_o is the initial number of VMs.

B. DU PC

The DU server serves as a crucial intermediary component connecting the CU server and the RU unit. This component encompasses several essential activities, namely the physical layer, media access control layer, and transport layer. The rationale behind the proximity of these functions to the users is to minimise call activation time by avoiding communications with distant cloud centres. The assumption was made that the DU server is also virtualized. Each VM is tasked with the responsibility of establishing communication with a single RU. In order to determine the PC of the DU server, we employed a similar methodology to that utilised for the CU server, as outlined below:

$$P_{DU}(t) = \frac{P_c}{\delta} + (P^o_{DU} - \frac{P_c}{\delta})e^{-\delta(t-t_o)} \tag{17}$$

and

$$P_{DU}(N) = \frac{P_c}{\delta} + (P^o_{DU} - \frac{P_c}{\delta})e^{-\delta(N-N_o)} \tag{18}$$

C. RU PC

The RU is connected to the UEs from one side, and to the DU from the other side. It was assumed the RU is a bare device and not virtualized. It includes two main units, the radio frequency and the power amplifier unit. The PC of the power amplifier can be evaluated by considering the maximum transmission power of the RU.

$$P_{PA} = \frac{P_{NB}}{\eta} \tag{19}$$

where η denotes the PA's efficiency. It is noted that this unit is affected by the number of transmitted RBs, if we substitute $P_{NB} = P_{rb} \times N_{rb}$, Equation 19, is written as:

$$P_{PA} = \frac{P_{rb} \times N_{rb}}{\eta} \tag{20}$$

This means the more bandwidth allocated to the system, the more PC of the power amplifier. In addition, the radio frequency unit P_{RF} is slightly affected by the transmitted RBs [57].

Hence, the radio unit PC P_{RU} is the summation of the radio frequency and power amplifier units.

$$P_{RU} = P_{RF} + P_{PA} \tag{21}$$

IV. QUANTUM PC

Modelling the entanglement-wise principal component is of utmost importance. In practical terms, a laser beam is employed to excite a nonlinear crystal known as a beta barium borate (BBO) crystal. The semiconductor laser induces the generation of photon pairs in the crystal, wherein the frequency of each twin photon is halved and the wavelength is doubled compared to the initial laser beam. The PC associated with the laser is represented as P_{laser}, while the PC connected to the detector is designated as P_{det}. Similarly, the PC linked to the driver is indicated as P_{driver}. The expression for the quantum entanglement's PC, denoted as P_{ent}, is provided as follows:

$$P_{ent} = P_{laser} + P_{det} + P_{driver} \tag{22}$$

In addition, the traditional PC is assumed as follows:

$$P_{traditional} = \frac{(P_{CU} + P_{DU})}{P_{loss}} + \frac{P_{RU}}{P_{AcDc}} \tag{23}$$

where P_{loss} denotes the PC of the AC-AC P_{AC}, AC-DC P_{DC} and cooling P_{cool}, where

$$P_{loss} = P_{AC} + P_{DC} + P_{cool} \tag{24}$$

The cooling PC is excluded in the RU unit, as.

$$P_{AcDc} = P_{AC} + P_{DC} \tag{25}$$

The quantum PC $P_{quantum}$ is given by adding P_{ent} to the quantum case while deducting the power saving value P_{save} from the traditional consumption, as follows:

$$P_{quantum} = (P_{traditional} - P_{save}) + P_{ent} \tag{26}$$

To formulate P_{save}, the PC of the messages from the server of interest to other servers is calculated. Since the number of bits in each signalling message is known parameter, the number of RBs in each message to be sent to server s; using modulation type x; can be evaluated as

$$N^{msg}_{rb,s,x} = \frac{L_{msg,s,x}}{L^{s,x}_{rb}} \tag{27}$$

where $L_{msg,s,x}$ denotes the number of bits contained in the message that is sent to server s using x type of modulation. To obtain the power contained in the message, $N^{msg}_{rb,s,x}$ is multiplied by the PC of RB, as follows:

$$P^{msg}_{rb,s,x} = N^{msg}_{rb,s,x} \times P_{rb} \tag{28}$$

Hence, P_{save} is the power consumed for round-trip transmitting the messages with $P^{msg}_{rb,s,x}$ to a total number of servers S, i.e.:

$$P_{save} = S \times P^{msg}_{rb,s,x} \tag{29}$$

V. SELECTING SPECIFIC SERVERS

According to [58], the process of generating entanglement is facilitated when a smaller number of photons are created. This phenomenon has the potential to impact the necessary optical and electrical equipment, as well as the acquisition of exceptionally pure photons, a high-fidelity system, and more prominent polarisation states. The current task necessitates the selection of servers to which these photons will be transmitted, in order to facilitate load balancing in a manner that minimises the amount of created photons. This contributes to the reduction in the quantity of photons created and mitigates the intricacy of the proposed system. The cloud infrastructure consists of a substantial quantity of servers, which can be partitioned into smaller subsets. Each subset can then be subjected to an optimisation procedure.

Suppose a cloud with servers set S, each server s is serving several users, where the total number of users is U, and $u_{s,n}$ is the users that is served by n-th VM in the server s. It is required to optimise the EE of these servers so as the servers with minimum EE are selected for the quantum load balancing. In general, the EE is given by $E = \frac{C}{PC}$ The power consumption, denoted as PC, can be represented by either $P_{traditional}$ or $P_{quantum}$ depending on the system being assessed. The selection method involves the utilisation of the traditional PC. Upon resolving the optimisation problem, the PC $P_{quantum}$ can be employed to facilitate the comparison between the pre-optimization and post-optimization states. The proposed optimization method have been implemented for single and two servers, following some assumptions regarding both cases, as follows:

1) The number of VMs that are allocated to each server is proportional to the number of users U. We assumed that $N_s = U_s/r$, where r is the share of the VMs is the server s. For example, if server 1 is connected to 4 UEs and server 2 is connected to 2 UEs and r=2, it means server 1 and 2 contain 2VMs and 1 VM, respectively. In addition, U_s denotes the total number of users within servers s; $U_s \in U$ and U is the total number of users.
2) The maximum consumption of a server is denoted as PCU_{max}.
3) The minimum achieved sum data rate is guaranteed via the C_{th} constraint, where C_{th} denotes the threshold data rate.
4) The PC model is reformulated as linear model to force the effect of the N_s and RBs variables within both the PC and the data rate.
5) In case of two servers, the linear PC model is also converted to linear model as this has same effect on both servers.

In a single server, the EE is formulated as the division of data rate measurement, i.e, $C = \sum_{u=1}^{U_s} B\ log(1 + \frac{Pt_{s,n,u}H_{s,n,u}}{N_0+I})$; and the measurement of PC, i.e, $PC = P_{CUs} + (N_s \times P_{vm}^{s,v,n})$. Whereas, $Pt_{s,n,u}$ and $H_{s,n,u}$ denote the power transmitted and channel gain of user u of VM n in the server s.

$$\max \quad \frac{\sum_{u=1}^{U_s} B\ log(1 + \frac{Pt_{s,n,u}H_{s,n,u}}{N_0+I})}{P_{CU_s} + (N_s \times P_{vm}^{s,v,n})} \tag{30}$$

$$\text{s.t.} \quad N_s \leq N \tag{31}$$

$$P_{CU_s} + (N_s \times P_{vm}^{s,v,n}) \leq PCU_{max} \tag{32}$$

$$C \geq C_{thr} \tag{33}$$

$$N_s, P_{CU_s}, C \geq 0 \tag{34}$$

The first constraint (32) showed that the Vms of the any server N_s not exceeding the total number of VMs in the network N. Constraint 33 shows that the total consumption of the server cannot overcome the maximum consumption, while the average data rate of the users is fixed to more than the threshold rate C_{thr} in constraint 34.

Regarding the optimisation variables, it was assumed that $C = x$, $N_s = y$, $P_{CUs} = z$, substituting $N_s = U_s/r$, and $P_{vm}^{s,v,n} = a$. The channel capacity of the UEs of VM (x) is multiplied by the number of VMs (y) to produce the total capacity. Hence, the optimisation problem becomes:

$$\max \quad F(x, y, z) = \frac{rxy}{z + (ay)} \tag{35}$$

$$\text{s.t.} \quad y \leq N \tag{36}$$

$$z + ay \leq PCU_{max} \tag{37}$$

$$x \geq C_{thr} \tag{38}$$

$$x, y, z \geq 0 \tag{39}$$

Note that the type of the problem is nonlinear and hence, Lagrange multiplier solution is used to obtain the optimal values that maximise the problem. In which, the general expression is given by $\mathcal{L} = F(x, y) - \lambda\big(g(x, y) - c\big)$, where $g(x, y)$ is the constraint functions. The full solution can be found in Appendix A.

$$\begin{aligned}\mathcal{L} &= \frac{rxy}{z + (ay)} + \lambda_1(N - y) + \lambda_2(C - C_{thr}) \\ &\quad + \lambda_3(PCU_{max} - z - y) \\ \partial L_x &= \frac{ry}{z + (ay)} + \lambda_2 \\ \partial L_y &= \frac{rx(z + ay) - (raxy)}{(z + (ay))^2} + \lambda_1 - \lambda_3 \\ \partial L_z &= \frac{-rxy}{(z + (ay))^2} + \lambda_3\end{aligned}$$

Solving these equations results the following:

$$\begin{aligned}\lambda_1 &= \frac{rx}{z + ay} \\ \lambda_2 &= \frac{-rxy}{(z + ay)^2} \\ \lambda_3 &= \frac{-ry}{z + ay}\end{aligned}$$

Or can be solved with respect to x,y,z, as follows:

$$\begin{aligned}x &= \frac{-z\lambda_3}{r + a\lambda_3} \leq C_{thr} \\ y &= \frac{-z\lambda_3}{r + a\lambda_3} \leq N \\ z &= \frac{y(\lambda_1 a - rx)}{rx + -\lambda_1} \leq PCU_{max} - ay\end{aligned}$$

Given the values of a, r, N, PCU_{max}, C_{thr}, the values of x,y,z can be obtained.

The convexity of a problem can be evaluated by evaluating the second derivatives of the objective function with respect to various variables. The negative value of the second derivative implies a concave shape, while the positive value of the second derivative implies convexity. Note that the second derivatives of the function with respect to x_1 and x_2 are equal to zero, which does not provide definitive conclusions on the function's convexity. On the other hand, the second derivative of other variables produces positive and negative values. Based on these observations, it can be concluded that the issue does not exhibit global convexity or concavity, as shown in Appendix C.

In the above formulation, the EE problem is presented to maximize the server's EE, while in the following formulation, the EE optimisation of two servers is presented:

$$\max \quad \frac{ry_1x_1}{z_1+ay_1}+\frac{ry_2x_2}{z_2+ay_2} \tag{40}$$

$$\text{s.t.} \quad x_1+x_2 \geq C_{thr} \tag{41}$$

$$y_1+y_2 \leq N \tag{42}$$

$$z_1+ay_1 \leq PCU_1 \tag{43}$$

$$z_2+ay_2 \leq PCU_2 \tag{44}$$

$$z_1+z_2 = PCU_{1,init}+PCU_{2,init} \tag{45}$$

$$\begin{aligned}&\mathcal{L}(x_1,x_2,y_1,y_2,z_1,z_2,\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5)\\&=\frac{r_1\cdot y_1\cdot x_1}{z_1+a_1\cdot y_1}+\frac{r_2\cdot y_2\cdot x_2}{z_2+a_2\cdot y_2}\\&\quad+\lambda_1\cdot(x_1+x_2-C_{\text{thr}})+\lambda_2\cdot(N-y_1-y_2)\\&\quad+\lambda_3\cdot(\text{PCU1}-z_1-a_1\cdot y_1)+\lambda_4\cdot(\text{PCU2}-z_2-a_2\cdot y_2)\\&\quad+\lambda_5\cdot(z_1+z_2-(\text{PCU1ini}+\text{PCU2init}))\end{aligned}$$

where

$$\begin{aligned}\partial L_{x1}&=\frac{r_1\cdot y_1}{z_1+a_1\cdot y_1}+\lambda_1=0\\\partial L_{x2}&=\frac{r_2\cdot y_2}{z_2+a_2\cdot y_2}+\lambda_1=0\\\partial L_{y1}&=\frac{r_1\cdot x_1\cdot(z_1+a_1\cdot y_1)-r_1\cdot y_1\cdot a_1\cdot x_1}{(z_1+a_1\cdot y_1)^2}\\&\quad-\lambda_2-\lambda_3\cdot a_1=0\\\partial L_{y2}&=\frac{r_2\cdot x_2\cdot(z_2+a_2\cdot y_2)-r_2\cdot y_2\cdot a_2\cdot x_2}{(z_2+a_2\cdot y_2)^2}\\&\quad-\lambda_2-\lambda_4\cdot a_2=0\\\partial L_{z1}&=-\frac{r_1\cdot y_1\cdot x_1}{(z_1+a_1\cdot y_1)^2}-\lambda_3+\lambda_5=0\\\partial L_{z2}&=-\frac{r_2\cdot y_2\cdot x_2}{(z_2+a_2\cdot y_2)^2}-\lambda_4+\lambda_5=0\end{aligned}$$

Subsequently, the solution of these equations can be found in Appendix B, where $\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5,x_1$, the optimisation variables are obtained. In addition, the proof of convexity/concavity can be found in Appendix C.

VI. ENTANGLEMENT INTEGRATION

A. ENTANGLEMENT RELIABILITY

The attainment of a fully consistent rate of entangled photon production presents difficulties stemming from inherent quantum uncertainty. Nonetheless, researchers have made notable advancements in enhancing the dependability and uniformity of this phenomenon. The utilization of sophisticated experimental configurations and meticulous control methods has facilitated the achievement of enhanced consistency in generation rates. The maintenance of a steady rate of entangled photon creation is of utmost importance for practical applications, such as quantum communication or quantum computing. A range of methodologies have been developed to enhance the reliability and predictability of entangled photon output. These strategies include advancements in laser source stability, improvements in crystal quality, and the development of more sophisticated triggering mechanisms. There are various strategies to accomplish this objective, such as:

1) Pulse-pumped sources enable the generation of entangled photon pairs at regular intervals through the utilisation of pulsed laser sources and precise timing control. This approach offers enhanced control and consistency in the production rate [59].
2) Triggered single-photon sources refer to devices that are capable of generating single photons upon request. These sources utilise methods such as spontaneous parametric down-conversion (SPDC) that can be triggered by external signals. Through this mechanism, researchers are able to control the emission of entangled photon pairs [60], [61].
3) The concept of cascaded entanglement involves the utilisation of several nonlinear crystals or quantum emitters in order to enhance the predictability of photon emission and achieve a more consistent rate of entangled photon pairs [62].
4) The refinement of experimental variables, including precise temperature regulation, crystal alignment, and improved photon detection capabilities, plays a crucial role in attaining a higher level of stability and control in the generation rate [63].

B. POWER REQUIRED FOR GENERATION AND RECEIVING

Using orders of mWs as output power, the pump laser power is used in the generation of the entanglement photons, regardless the number of generated entangled pairs [64]. The power consumed of the laser source is in the range of a few watts, as shown below:

- **200 mW to 500 mW output power**: Consumes power ranging from approximately 3.5 watts to 7 watts during operation [65].
- **1700 mW output power**: Consumes higher power, typically around 1.7 watts [66].
- **3500 mW output power**: Requires power around 12 watts during operation [67].

However, in the receiving side, the detectors usually require the same amount of power to detect the very low-power signals. With less than 1W, as mentioned in [68] and up to 6 Watts as mentioned in [69]. However, the total transmitting and receiving process of entangled photons is in the range of 10 W. This amount, however, can be overcome by the gain of the proposed method that suggested mitigating up to half the power consumption of the server. The latter consumes power in the range of 118W to 365W, as mentioned in [70]. Hence, in the worst case scenario, even if the gain of the proposed method has achieved only 10 % server power reduction, the gain of the power will be between 18 and 36.5 W.

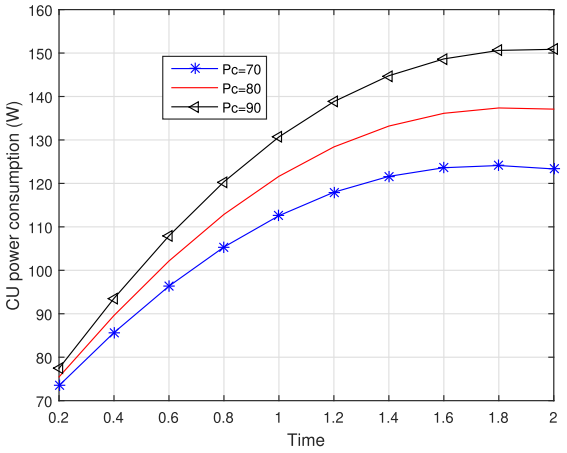

FIGURE 3. Effect of P_c on the PC of the CU server.

VII. RESULTS AND DISCUSSION

In the given optimisation problem, the objective is to maximise a utility function that represents the effectiveness of allocating resources amongst single or two servers and the users. The system comprises many parameters, as shown in Table 1, as follows:

- N_s: Number of VMs allocated to server s.
- P_{CU_s}: Power consumption of central unit.
- C: Capacity of the system.
- U_s: Total number of users served by server s.
- B: Bandwidth of the system.
- $Pt_{s,n,u}$: Transmit power from server s to user u via VM n.
- $H_{s,n,u}$: Channel gain from server s to user u via VM n.
- N_0: Noise power spectral density.
- I: Interference.
- PCU_{max}: Maximum power consumption of the central unit.
- $P_{vm}^{s,v,n}$: Power consumption of VM on server s.

Fig. 3 shows the PC comparison of the CU server with respect to different values of the P_c, while considering P_{CU}^o is 60W, $t_o = 0$, and t is up to 2 seconds. The model shows that the more consumed power P_c, the steady state of the model becomes ideal due to the overcoming behaviour of P_c value over the static value. In Addition, δ has been taken as 0.4. However, these values can be adjusted based on the different characteristic of the servers. Meaning, given the device's specifications, the maximum and static consummations' can be allocated in the model to present the mediate PC values with respect to the time or number of virtual machines.

In the next Figure, the P_c is fixed to 90W, while the values of the P_{CU}^o are changed. It is shown that the value of the static consumption only affects the total value of the consumption. The other assumption are kept the same as Fig 3.

The following Fig. 5 shows the effect of VMs on the PC of the server. Note that such server might be considered as CU or DU server. The values of P_c and P_{CU}^o are fixed to 80W and 70W, respectively. This figure shows the different behaviours of the model with respect to different values of α, which reflects on different servers specifications.

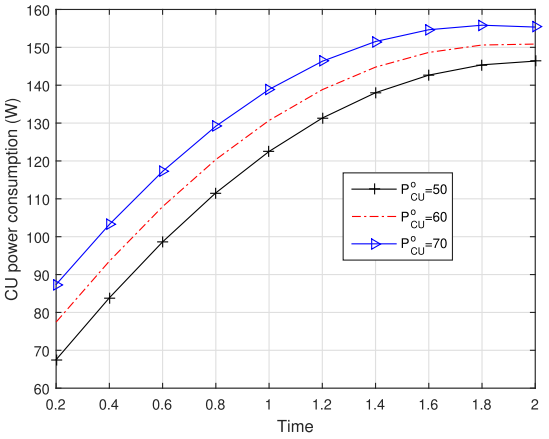

FIGURE 4. Effect of P_{CU}^o on the PC of the CU server.

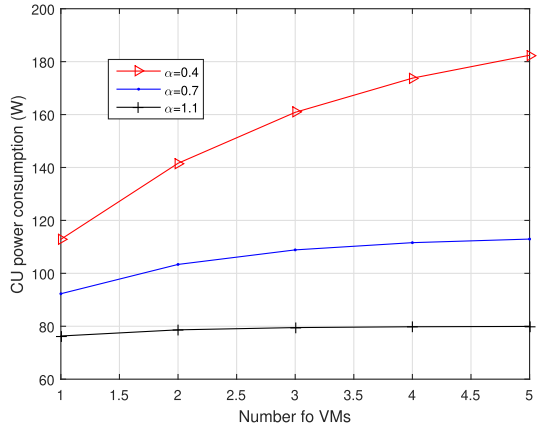

FIGURE 5. Effect of number of virtual machines on the PC of the CU server using different values of α.

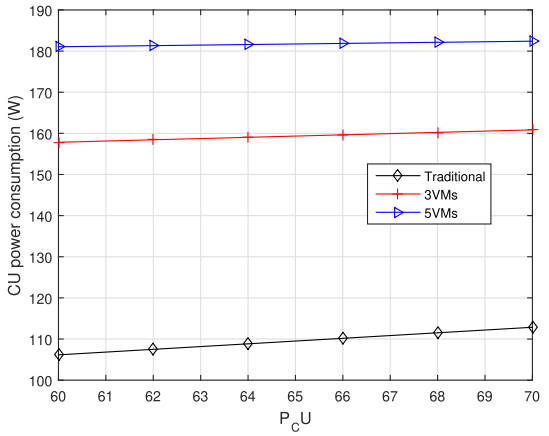

FIGURE 6. Effect of number of virtual machines on the PC of the CU server with $\alpha = 0.4$.

Subsequently, this study compares the PC of traditional server with virtualized server, assuming different number of VMs. The result depicted that traditional server can maintain low power consumption compared to virtualized server, as shown in Fig 6.

Nevertheless, this comparison lacks realism since the virtualized server with 3VMs effectively replaces 3 operating servers, while a server with 5VMs replaces 5 traditional servers, and so on. Consequently, an alternative comparison

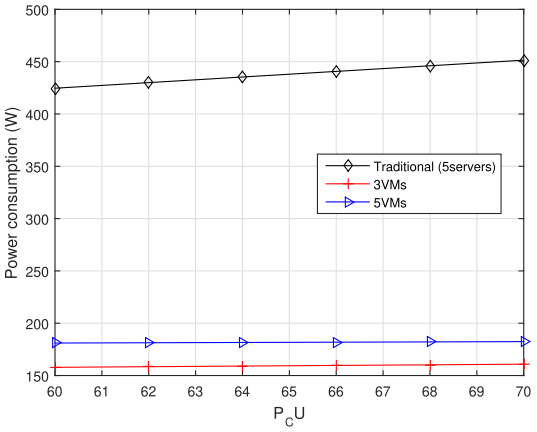

FIGURE 7. Effect of number of virtual machines on the PC of the network with $\alpha = 0.4$, operating 5 servers within the traditional network.

TABLE 1. Model parameters.

Factor	Value	Unit	Factor	Value	Unit
P_{bit}	0.3	W	L_{bit}	20	bits
N_{rb}	2	RBs	C_{thr}	100	bps
P^o_{CU1}	70	W	P^o_{CU2}	70	W
$PCU1$	100	W	$PCU2$	110	W
r_1	0.6	-	r_2	0.4	W
P_{driver}	1	W	P_{synch}	1	W
P_{delta}	0.4	-	P	1	W
B	20	MHz	P^o_{CU}	60	W

was conducted to accurately assess power usage at the system level. Fig 7. shows the PC comparison of the traditional network, containing 5 servers and virtualised servers operating with 3VMs and 5VMs, respectively. It is clear that the virtualised network scores less PC than the traditional counterpart.

To linearize and simplify the model, we have assumed the model contains the static and dynamic consumption, such as $P^o_{CU,s} + (N_s \times P^{s,v,n}_{vm})$. Not to forget it is crucial to decide the number of VMs to be placed in one server based on the coming load, which is also an optimization problem that contributes to lessen the power consumed by each server. However, in this work, to hold the objective function being dependant on the number of VMs in both the channel capacity and the PC. We have assumed that the number of VMs in the server is based on its connected users and a weighting parameter r, that is decidable and agreed amongst the network operators based on their ORAN policy agreement. Subsequently, when the value of r is decided, it affects the value of VMs, which affects the value of E. Furthermore, the optimisation problem is further constrained by the involvement of many vendors and operators. This is because the sharing policy might either involve complete sharing of cloud resources or be based on a weighing technique. In either case, it is necessary to develop new optimisation algorithms that prioritise the following issues:

1) The cloud is shared by network suppliers or operators. The remedy to the problem can vary depending on the individual's specific settings of PC for their devices.
2) The network operator allocates a portion of the cloud resources, either small or large, depending on its projected user base or through a pay-as-you-go approach. This implies that the values of PCs undergo temporal fluctuations.
3) In the ORAN architecture, many operators cohabit within a shared server environment. The current inquiry pertains to whether all of these operators are in consensus over the utilisation of virtualization technology to facilitate connectivity for their subscribers. In this scenario, it is anticipated that there will be an increase in latency. Therefore, the inclusion of a delay restriction can be incorporated into the optimisation problem.

The 'fmincon' solver has been employed as the preferred optimisation method due to its ability to effectively handle nonlinear objective functions and constraints, encompassing both inequality and equality constraints. The active-set approach and sequential quadratic programming (SQP) are two algorithms that are utilised and compared. The SQP method is a widely used approach for addressing nonlinear constrained optimisation issues. The method employed is iterative in nature, wherein the subproblem is resolved within the confines of the feasible zone. Simultaneously, the overarching problem is progressively approximated through the utilisation of a quadratic model. This method is well recognised as a highly efficient and effective approach for addressing complex nonlinear optimisation problems that involve both equality and inequality requirements.

The SQP algorithm integrates components from both Newton's approach, which involves second-order optimisation techniques, and the Lagrange multipliers method, which addresses the handling of constraints. The algorithm commences with initialising and selecting an initial feasible point $x(0)$, followed by setting the iteration counter k to zero. Next, we will proceed with the construction of a quadratic approximation for both the objective function and the constraint, centred around the point denoted as $x(k)$. Next, the quadratic sub-problem is solved in order to choose a search direction $p(k)$ that minimises the objective function. Subsequently, the point can be updated by employing the formula $x(k+1) = x(k) + \gamma p(k)$, where γ represents a step size that ensures both feasibility and progress towards the optimal value. Subsequently, the convergence can be assessed by employing specified criteria, such as evaluating constraint violation or monitoring changes in the objective function. Lastly, the counter k is incremented by adding 1, denoted as $k = k + 1$, and the process returns to step 2.

In contrast, the active set approach employs the identical criteria as the SQP method, however with convergence achieved when the solution to the sub-problem satisfies the Karush-Kuhn-Tucker (KKT) conditions. The two methods have been compared in terms of their performance using the given parameters. Figure 8 shows the different decision variables with respect to the objective function values of the SQP.

In Figure 9, the decision variables were depicted in relation to the objective function when employing the Active-set algorithm.

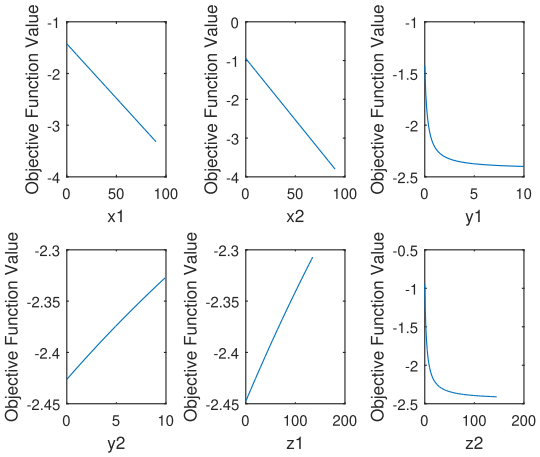

FIGURE 8. Decision parameters with respect to the objective value using SQP algorithm.

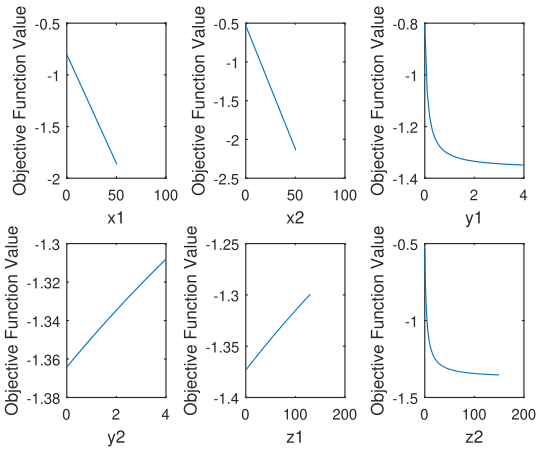

FIGURE 9. Decision parameters with respect to the objective value using Active-set algorithm.

In general, the SQP algorithm out perform the Active-set method. A comparison has been made between the two methods while adjusting some of the decision variables, such as C_{thr}, as shown in Fig. 10, when considering 20 users.

Note that SQP algorithm is superior to the Active Set method in achieving higher objective function values because SQP is well-suited for handling nonlinear constraints, which are common in complex systems like ORAN. Its ability to efficiently handle nonlinear constraints allows it to explore a broader solution space and identify more optimal solutions. SQP's global optimization capabilities enable it to search for solutions that yield higher EE values across diverse scenarios, leading to superior performance compared to the Active Set method. The SQP algorithm's iterative process allows for faster convergence, enabling it to explore the solution space more thoroughly and identify higher EE solutions within a shorter timeframe. Additionally, SQP's robustness to initialization makes it less sensitive to initial guess or starting point, increasing the likelihood of discovering higher EE solutions regardless of initial conditions.

The convergence characteristics of the optimisation algorithm can be impacted by a range of aspects, encompassing the starting solution, the selection of the optimisation

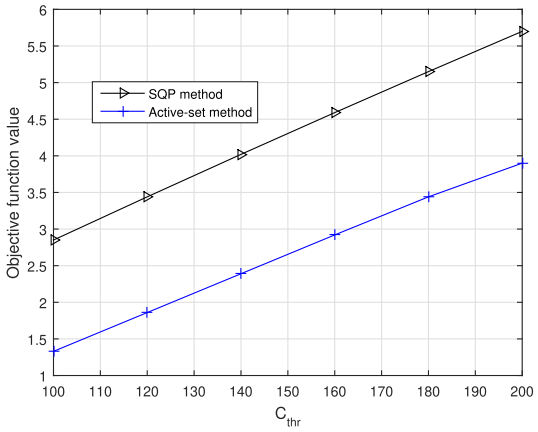

FIGURE 10. The threshold data rate with respect to the objective value using SQP and Active-set algorithms.

strategy, the attributes of the problem under consideration, and the particular configurations of the optimisation approach. Certain algorithms may demonstrate rapid convergence, while others may require a larger number of iterations to achieve a satisfactory outcome. In the real implementation, the suggested comparison may not produce favourable outcomes within the specific context of the requested work. Irrespective of the specific method utilised, it is feasible to ascertain the server exhibiting the lowest or maximum energy efficiency (EE), albeit with differing numbers for each approach. While heuristics can provide alternative approaches for comparing algorithms, the selection criteria for servers does not necessitate comprehensive knowledge of which method yields higher energy efficiency. Rather, it requires selecting certain servers. Therefore, conducting more comparisons is only time-consuming.

VIII. CONCLUSION AND FUTURE TRENDS

The installation of a substantial quantity of virtual machines, wherein each virtual machine consumes processing resources, leads to an expansion regarding the PC. Therefore, it is imperative to optimise the allocation of VMs across several operators in order to provide a highly efficient network. This study presents a power model that examines the impact of various network factors, notably the time factor and the quantity of VMs, on power consumption. This model is considered new because to its focus on the ORAN architecture. Furthermore, it is characterised by its simplified and realistic approach, distinguishing it from other models discussed in existing literature, which may involve complex functions, hardware, or software components. This model employed in the EE maximisation problem incorporated the proposed quantum entanglement approach. The process involves a comparison of the servers in order to determine the most suitable load balancing servers from the available options within the ORAN cloud. The optimisation problem was solved with the Lagrange multiplier technique and afterwards numerically solved employing the SQP approach, which demonstrated superior performance compared to the Active-set methods by about 45%.

In the future, the coexistence of several network operators within ORAN servers necessitates the optimisation of resource allocation to determine which operator can derive the greatest advantage from the servers' resources at a certain time window. The determination of various weights, such as bandwidth and power, is contingent upon the cost factor, which is ultimately influenced by the concept of infrastructure as a service.

APPENDIX A

Solving the first objective function with respect to x, y, z.

Given equations:

1. $(ry)/(z+ay)+\lambda_3=0$
2. $(rx(z+ay)-rxya)/((z+ay)^2)-\lambda_1-\lambda_2 a=0$
3. $-(rxy)/((z+ay)^2)-\lambda_2=0$

Equation 1:

$$(ry)/(z+ay)+\lambda_3=0$$

Multiplying both sides by $(z+ay)$:

$$ry+\lambda_3(z+ay)=0$$

Expanding:

$$ry+3lambda_3 z+\lambda_3 ay=0$$

Isolating y:

$$y(r+\lambda_3 a)=-\lambda_3 z$$

Dividing by $(r+\lambda_2 a)$:

$$y=-(\lambda_3 z)/(r+\lambda_3 a)$$

Equation 2:

$$(rx(z+ay)-rxya)/((z+ay)^2)-\lambda_1-\lambda_2 a=0$$

Multiplying both sides by $((z+ay)^2)$:

$$rx(z+ay)-rxya-\lambda_1((z+ay)^2)-\lambda_2 a((z+ay)^2)=0$$

Expanding:

$$rxz+rxay-rxya-\lambda_1(z^2+2ayz+a^2y^2)-\lambda_2 a(z^2+2ayz+a^2y^2)=0$$

Collecting for x:

$$rxz-rxya=\lambda_1(z^2+2ayz+a^2y^2)+\lambda_2 a(z^2+2ayz+a^2y^2)-rxay$$

Isolating x:

$$rx(z-ay)=(\lambda_1+\lambda_2 a)*(z^2+2ayz+a^2y^2)$$

Dividing by $(z-ay)$:

$$x=((\lambda_1+\lambda_2 a)(z^2+2ayz+a^2y^2))/(r(z-ay))$$

Equation 3:

$$-(rxy)/((z+ay)^2)-\lambda_2=0$$

Multiplying both by $((z+ay)^2)$:

$$-rxy-\lambda_2((z+ay)^2)=0$$

Expanding:

$$-rxy-\lambda_2(z^2+2ayz+a^2y^2)=0$$

Isolating y:

$$y(-rx-\lambda_2 a^2)=-\lambda_2(z^2+2ayz)$$

Dividing by $(-rx-\lambda_2 a^2)$:

$$y=-(\lambda_2(z^2+2ayz))/(-rx-\lambda_2 a^2)$$

Hence, the solutions for x and y are as follows:

$$x=((\lambda_1+\lambda_2 a)(z^2+2ayz+a^2y^2))/(r(z-ay))$$
$$y=-(\lambda_3 z)/(r+\lambda_3 a)$$

Solving for z

1. First equation:

$$\lambda_3=-(ry)/(z+ay)$$

2. Second equation:

$$(rx(z+ay)-rxya)/((z+ay)^2)=\lambda_1+\lambda_2 a$$

3. Third equation:

$$\lambda_2=-(rxy)/((z+ay)^2)$$

Eliminating λ_2 from Eqs. (2) and (3):

$$(rx(z+ay)-rxya)/((z+ay)^2)=\lambda_1-(rxy)/((z+ay)^2)$$

Cross-multiplying to ignore the denominators:

$$(rx(z+ay)-rxya)=\lambda_1(z+ay)-(rxy)$$

Distributing λ_1:

$$rxz+rxay-rxya=\lambda_1 z+\lambda_1 ay-rxy$$

Rearranging and isolating z:

$$rxz-\lambda_1 z=-rxay+\lambda_1 ay$$

Factoring out z:

$$z(rx-\lambda_1)=y(\lambda_1 a-rx)$$
$$z=(y(\lambda_1 a-rx))/(rx-\lambda_1)$$

APPENDIX B

Solving for the second objective function, solving for x_1, x_2, y_1, y_2, z_1, and z_2, the variables are isolated in each equation and then solving the resulting equations simultaneously, as follows:

1. First equation:

$$\frac{r_1 y_1}{z_1+a_1 y_1}+\lambda_1=0$$

Solving for y_1:

$$y_1=-\frac{\lambda_1(z_1+a_1 y_1)}{r_1}$$

Rearranging and isolating y_1:

$$(1+\frac{\lambda_1 a_1}{r_1})y_1 = -\frac{\lambda_1 z_1}{r_1}$$
$$y_1 = -\frac{\lambda_1 z_1}{r_1(1+\frac{\lambda_1 a_1}{r_1})}$$

2. Second equation:

$$\frac{r_2 y_2}{z_2 + a_2 y_2} + \lambda_1 = 0$$

Solve for y_2:

$$y_2 = -\frac{\lambda_1(z_2 + a_2 y_2)}{r_2}$$

Rearranging and isolating y_2:

$$(1+\frac{\lambda_1 \cdot a_2}{r_2})y_2 = -\frac{\lambda_1 z_2}{r_2}$$
$$y_2 = -\frac{\lambda_1 \cdot z_2}{r_2(1+\frac{\lambda_1 \cdot a_2}{r_2})}$$

3. Third equation:

$$\frac{r_1 x_1(z_1 + a_1 y_1) - r_1 y_1 a_1 x_1}{(z_1 + a_1 y_1)^2} - \lambda_2 - \lambda_3 a_1 = 0$$

Solving for x_1:

$$x_1 = \frac{r_1 y_1 a_1 x_1 - r_1 x_1 z_1 + \lambda_2(z_1 + a_1 y_1)^2}{r_1 y_1 a_1}$$

Rearranging and isolating x_1:

$$(1-\frac{\lambda_2}{r_1 y_1 a_1})x_1 = \frac{\lambda_2(z_1 + a_1 y_1)^2 - r_1 x_1 z_1}{r_1 y_1 a_1}$$
$$x_1 = \frac{\lambda_2(z_1 + a_1 y_1)^2 - r_1 x_1 z_1}{r_1 y_1 a_1(1-\frac{\lambda_2}{r_1 y_1 a_1})}$$

4. Fourth equation:

$$\frac{r_2 x_2(z_2 + a_2 y_2) - r_2 y_2 a_2 x_2}{(z_2 + a_2 y_2)^2} - \lambda_2 - \lambda_4 a_2 = 0$$

Solving for x_2:

$$x_2 = \frac{r_2 y_2 a_2 x_2 - r_2 x_2 z_2 + \lambda_2(z_2 + a_2 y_2)^2}{r_2 y_2 a_2}$$

Rearranging and isolating x_2:

$$(1-\frac{\lambda_2}{r_2 y_2 a_2})x_2 = \frac{\lambda_2(z_2 + a_2 y_2)^2 - r_2 x_2 z_2}{r_2 y_2 a_2}$$
$$x_2 = \frac{\lambda_2(z_2 + a_2 y_2)^2 - r_2 x_2 z_2}{r_2 y_2 a_2(1-\frac{\lambda_2}{r_2 y_2 a_2})}$$

5. Fifth equation:

$$-\frac{r_1 y_1 x_1}{(z_1 + a_1 y_1)^2} - \lambda_3 + \lambda_5 = 0$$

Solve for z_1:

$$z_1 = \frac{r_1 x_1 y_1}{(1-\frac{\lambda_3}{r_1 y_1})}$$

6. Sixth equation:

$$-\frac{r_2 y_2 x_2}{(z_2 + a_2 y_2)^2} - \lambda_4 + \lambda_5 = 0$$

Solving for z_2:

$$z_2 = \frac{r_2 x_2 y_2}{(1-\frac{\lambda_4}{r_2 y_2})}$$

APPENDIX C

In order to examine the convexity of the propsed problem, it is necessary to evaluate the second derivatives of the objective function with respect to the variables x_1, x_2, y_1, y_2, z_1, and z_2.

1-The second derivative of the function f with respect to x_1 is equal to zero.

2-The second derivative of the function f with respect to x_2 is equal to zero. Given that the second derivative remains constant and non-negative, this does not yield definitive conclusions on the convexity of the function.

3-The second derivative of the function f with respect to y_1 can be expressed as follows:

$$\frac{\partial^2 f}{\partial y_1^2} = -2 r_1 x_1(z_1 + a_1 y_1)$$

The negative value of the second derivative implies a concave.

4-The second derivative of the function f with respect to y_2 can be expressed as follows:

$$\frac{\partial^2 f}{\partial y_2^2} = -2 r_2 x_2(z_2 + a_2 y_2)$$

The negative value of the second derivative suggests a concave shape.

5-The second derivative of the function f with respect to z_1 can be expressed as follows:

$$\frac{\partial^2 f}{\partial z_1^2} = 2\, r_1 y_1 x_1(2 a_1 y_1 - z_1)$$

6-The second derivative of the function f with respect to z_2 is given by:

$$\frac{\partial^2 f}{\partial z_2^2} = \frac{2 r_2 y_2 x_2(2 a_2 y_2 - z_2)}{z_2}$$

The positive value of the second derivative implies the presence of convexity.

Based on the observed indications of the second derivatives, it may be deduced that the issue does not exhibit global convexity or concavity.

REFERENCES

[1] J. X. Salvat, J. A. Ayala-Romero, L. Zanzi, A. Garcia-Saavedra, and X. Costa-Perez, "Open radio access networks (O-RAN) experimentation platform: Design and datasets," *IEEE Commun. Mag.*, vol. 61, no. 9, pp. 138–144, Sep. 2023.

[2] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 957–975, 2020.

[3] A. Masaracchia, V. Sharma, M. Fahim, O. A. Dobre, and T. Q. Duong, "Digital twin for open RAN: Towards intelligent and resilient 6G radio access networks," *IEEE Commun. Mag.*, vol. 61, no. 11, pp. 112–118, Nov. 2023.

[4] S. K. Singh, R. Singh, and B. Kumbhani, "The evolution of radio access network towards open-RAN: Challenges and opportunities," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Apr. 2020, pp. 1–6.

[5] A. Giannopoulos, S. Spantideas, N. Kapsalis, P. Gkonis, L. Sarakis, C. Capsalis, M. Vecchio, and P. Trakadas, "Supporting intelligence in disaggregated open radio access networks: Architectural principles, AI/ML workflow, and use cases," *IEEE Access*, vol. 10, pp. 39580–39595, 2022.

[6] A. S. Abdalla, P. S. Upadhyaya, V. K. Shah, and V. Marojevic, "Toward next generation open radio access networks: What O-RAN can and cannot do!" *IEEE Netw.*, vol. 36, no. 6, pp. 206–213, Nov. 2022.

[7] D. Mimran, R. Bitton, Y. Kfir, E. Klevansky, O. Brodt, H. Lehmann, Y. Elovici, and A. Shabtai, "Security of open radio access networks," *Comput. Secur.*, vol. 122, Nov. 2022, Art. no. 102890.

[8] Y. Cao, S.-Y. Lien, Y.-C. Liang, K.-C. Chen, and X. Shen, "User access control in open radio access networks: A federated deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 3721–3736, Jun. 2022.

[9] C. D. Alwis, A. Kalla, Q.-V. Pham, P. Kumar, K. Dev, W.-J. Hwang, and M. Liyanage, "Survey on 6G frontiers: Trends, applications, requirements, technologies and future research," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 836–886, 2021.

[10] V.-D. Nguyen, T. X. Vu, N. Thanh Nguyen, D. C. Nguyen, M. Juntti, N. Cong Luong, D. Thai Hoang, D. N. Nguyen, and S. Chatzinotas, "Network-aided intelligent traffic steering in 6G O-RAN: A multi-layer optimization framework," 2023, *arXiv:2302.02711*.

[11] N. Kazemifard and V. Shah-Mansouri, "Minimum delay function placement and resource allocation for open RAN (O-RAN) 5G networks," *Comput. Netw.*, vol. 188, Apr. 2021, Art. no. 107809.

[12] 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 Trans. Green Commun. Netw.*, vol. 2, no. 4, pp. 992–1001, Dec. 2018.

[13] J. Choi, S. Oh, and J. Kim, "Energy-efficient cluster head selection via quantum approximate optimization," *Electronics*, vol. 9, no. 10, p. 1669, Oct. 2020.

[14] M. Erhard, M. Krenn, and A. Zeilinger, "Advances in high-dimensional quantum entanglement," *Nature Rev. Phys.*, vol. 2, no. 7, pp. 365–381, Jun. 2020.

[15] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller, "Quantum repeaters: The role of imperfect local operations in quantum communication," *Phys. Rev. Lett.*, vol. 81, no. 26, pp. 5932–5935, Dec. 1998.

[16] M. Aspelmeyer, T. Jennewein, M. Pfennigbauer, W. R. Leeb, and A. Zeilinger, "Long-distance quantum communication with entangled photons using satellites," *IEEE J. Sel. Topics Quantum Electron.*, vol. 9, no. 6, pp. 1541–1551, Nov. 2003.

[17] L. Gyongyosi and S. Imre, "Distributed quantum computation for near-term quantum environments," *Proc. SPIE*, vol. 11726, Apr. 2021, Art. no. 117260I.

[18] I. P. Christov, "Spatial entanglement of fermions in one-dimensional quantum dots," *Entropy*, vol. 23, no. 7, p. 868, Jul. 2021.

[19] M. Chehimi and W. Saad, "Entanglement rate optimization in heterogeneous quantum communication networks," 2021, *arXiv:2105.14507*.

[20] C.-Z. Peng, T. Yang, X.-H. Bao, J. Zhang, X.-M. Jin, F.-Y. Feng, B. Yang, J. Yang, J. Yin, Q. Zhang, N. Li, B.-L. Tian, and J.-W. Pan, "Experimental free-space distribution of entangled photon pairs over 13 km: Towards satellite-based global quantum communication," *Phys. Rev. Lett.*, vol. 94, no. 15, Apr. 2005, Art. no. 150501.

[21] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, "Classical and quantum communication without a shared reference frame," *Phys. Rev. Lett.*, vol. 91, no. 2, Jul. 2003, Art. no. 027901.

[22] B. Da Lio, D. Bacco, D. Cozzolino, N. Biagi, T. N. Arge, E. Larsen, K. Rottwitt, Y. Ding, A. Zavatta, and L. K. Oxenløwe, "Stable transmission of high-dimensional quantum states over a 2-km multicore fiber," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 4, pp. 1–8, Jul. 2020.

[23] D. Attanayaka, P. Porambage, M. Liyanage, and M. Ylianttila, "Peer-to-peer federated learning based anomaly detection for open radio access networks," in *Proc. IEEE Int. Conf. Commun.*, May 2023, pp. 5464–5470.

[24] P. K. Thiruvasagam, V. Venkataram, V. R. Ilangovan, M. Perapalla, R. Payyanur, and V. Kumar, "Open RAN: Evolution of architecture, deployment aspects, and future directions," 2023, *arXiv:2301.06713*.

[25] I. Oussakel, P. Owezarski, P. Berthou, and L. Houssin, "Toward radio access network slicing enforcement in multi-cell 5G system," *J. Netw. Syst. Manag.*, vol. 31, no. 1, p. 8, Jan. 2023.

[26] S. Mondal and M. Ruffini, "Fairness guaranteed and auction-based X-haul and cloud resource allocation in multi-tenant O-RANs," 2023, *arXiv:2301.00597*.

[27] H.-M. Yoo, J.-S. Rhee, S.-Y. Bang, and E.-K. Hong, "Load balancing algorithm running on open ran ric," in *Proc. 13th Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, 2022, pp. 1226–1228.

[28] M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang, "Recent advances in cloud radio access networks: System architectures, key techniques, and open issues," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 2282–2308, 3rd Quart., 2016.

[29] A. Arnaz, J. Lipman, M. Abolhasan, and M. Hiltunen, "Toward integrating intelligence and programmability in open radio access networks: A comprehensive survey," *IEEE Access*, vol. 10, pp. 67747–67770, 2022.

[30] Y. Ye, T. Zhang, and L. Yang, "Joint user association and resource allocation for load balancing in RAN slicing," *Phys. Commun.*, vol. 49, Dec. 2021, Art. no. 101459.

[31] E. Amiri, N. Wang, M. Shojafar, and R. Tafazolli, "Optimizing virtual network function splitting in open-RAN environments," in *Proc. IEEE 47th Conf. Local Comput. Netw. (LCN)*, Sep. 2022, pp. 422–429.

[32] C.-H. Lai, L.-H. Shen, and K.-T. Feng, "Intelligent load balancing and resource allocation in O-RAN: A multi-agent multi-armed bandit approach," 2023, *arXiv:2303.14355*.

[33] O. Orhan, V. N. Swamy, T. Tetzlaff, M. Nassar, H. Nikopour, and S. Talwar, "Connection management xAPP for O-RAN RIC: A graph neural network and reinforcement learning approach," in *Proc. 20th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2021, pp. 936–941.

[34] N. Budhdev, A. Maity, M. C. Chan, and T. Mitra, "Load balancing for a user-level virtualized 5G cloud-RAN," in *Proc. 17th ACM Workshop Mobility Evolving Internet Archit.*, Oct. 2022, pp. 1–6.

[35] E. A. Ramos da Paixão, R. F. Vieira, W. V. Araújo, and D. L. Cardoso, "Optimized load balancing by dynamic BBU-RRH mapping in C-RAN architecture," in *Proc. 3rd Int. Conf. Fog Mobile Edge Comput. (FMEC)*, Apr. 2018, pp. 100–104.

[36] B. Mahapatra, R. Kumar, A. K. Turuk, and S. K. Patra, "CLB: A multilevel co-operative load balancing algorithm for C-RAN architecture," *Digit. Commun. Netw.*, vol. 5, no. 4, pp. 308–316, Nov. 2019.

[37] D. Sharma, S. Singhal, A. Rai, and A. Singh, "Analysis of power consumption in standalone 5G network and enhancement in energy efficiency using a novel routing protocol," *Sustain. Energy, Grids Netw.*, vol. 26, Jun. 2021, Art. no. 100427.

[38] T. Pamuklu, S. Mollahasani, and M. Erol-Kantarci, "Energy-efficient and delay-guaranteed joint resource allocation and DU selection in O-RAN," in *Proc. IEEE 4th 5G World Forum (5GWF)*, Oct. 2021, pp. 99–104.

[39] M. Dayarathna, Y. Wen, and R. Fan, "Data center energy consumption modeling: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 732–794, 1st Quart., 2016.

[40] R. S. Alhumaima and H. S. Al-Raweshidy, "Modelling the power consumption and trade-offs of virtualised cloud radio access networks," *IET Commun.*, vol. 11, no. 7, pp. 1158–1164, May 2017.

[41] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[42] H. Holtkamp, G. Auer, V. Giannini, and H. Haas, "A parameterized base station power model," *IEEE Commun. Lett.*, vol. 17, no. 11, pp. 2033–2035, Nov. 2013.

[43] R. Shea, H. Wang, and J. Liu, "Power consumption of virtual machines with network transactions: Measurement and improvements," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, Apr. 2014, pp. 1051–1059.

[44] Q. Huang, F. Gao, R. Wang, and Z. Qi, "Power consumption of virtual machine live migration in clouds," in *Proc. 3rd Int. Conf. Commun. Mobile Comput.*, Apr. 2011, pp. 122–125.

[45] M. Marcu and D. Tudor, "Power consumption measurements of virtual machines," in *Proc. 6th IEEE Int. Symp. Appl. Comput. Intell. Informat. (SACI)*, May 2011, pp. 445–449.

[46] Y. Al-Karawi, R. S. Alhumaima, and H. Al-Raweshidy, "Quality of service of quantum entanglement in mobile networks," *IEEE Access*, vol. 9, pp. 167242–167251, 2021.
[47] T. Darras, B. E. Asenbeck, G. Guccione, A. Cavaillès, H. Le Jeannic, and J. Laurat, "A quantum-bit encoding converter," *Nature Photon.*, vol. 17, no. 2, pp. 165–170, Feb. 2023.
[48] F. Benatti, R. Floreanini, F. Franchini, and U. Marzolino, "Entanglement in indistinguishable particle systems," *Phys. Rep.*, vol. 878, pp. 1–27, Sep. 2020.
[49] M. H. Appel, A. Tiranov, S. Pabst, M. L. Chan, C. Starup, Y. Wang, L. Midolo, K. Tiurev, S. Scholz, A. D. Wieck, A. Ludwig, A. S. Sørensen, and P. Lodahl, "Entangling a hole spin with a time-bin photon: A waveguide approach for quantum dot sources of multiphoton entanglement," *Phys. Rev. Lett.*, vol. 128, no. 23, Jun. 2022, Art. no. 233602.
[50] P. S. Aithal, "Advances and new research opportunities in quantum computing technology by integrating it with other ICCT underlying technologies," *Int. J. Case Stud. Bus., IT, Educ.*, vol. 7, no. 3, pp. 314–358, Sep. 2023.
[51] F. Ahmed and P. Mähönen, "Quantum computing for artificial intelligence based mobile network optimization," in *Proc. IEEE 32nd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2021, pp. 1128–1133.
[52] M. I. Habibie, J. Hamie, and C. Goursaud, "A performance comparison of classical and quantum algorithm for active user detection," in *Proc. IEEE 23rd Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2022, pp. 1–5.
[53] M. H. Adnan, Z. Ahmad Zukarnain, and N. Z. Harun, "Quantum key distribution for 5G networks: A review, state of art and future directions," *Future Internet*, vol. 14, no. 3, p. 73, Feb. 2022.
[54] T. Q. Duong, J. A. Ansere, B. Narottama, V. Sharma, O. A. Dobre, and H. Shin, "Quantum-inspired machine learning for 6G: Fundamentals, security, resource allocations, challenges, and future research directions," *IEEE Open J. Veh. Technol.*, vol. 3, pp. 375–387, 2022.
[55] R. S. Alhumaima, S. T. Alwan, and R. K. Ahmed, "Mitigating X2-AP interface cost using quantum teleportation," *IET Netw.*, vol. 9, no. 5, pp. 247–254, Sep. 2020.
[56] C. Pang, A. Hindle, B. Adams, and A. E. Hassan, "What do programmers know about software energy consumption?" *IEEE Softw.*, vol. 33, no. 3, pp. 83–89, May 2016.
[57] M. Tayyab, G. P. Koudouridis, X. Gelabert, and R. Jäntti, "Signaling overhead and power consumption during handover in LTE," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–6.
[58] M. B. Plenio, J. Hartley, and J. Eisert, "Dynamics and manipulation of entanglement in coupled harmonic systems with many degrees of freedom," *New J. Phys.*, vol. 6, p. 36, Mar. 2004.
[59] M. Marghany, *Synthetic Aperture Radar Imaging Mechanism for Oil Spills*. Gulf Professional, 2019.
[60] Y. Li, "Methods of generating entangled photon pairs," *J. Phys., Conf.*, vol. 1634, no. 1, Sep. 2020, Art. no. 012172.
[61] H. C. Lim, A. Yoshizawa, H. Tsuchida, and K. Kikuchi, "Stable source of high quality telecom-band polarization-entangled photon-pairs based on a single, pulse-pumped, short PPLN waveguide," *Opt. Exp.*, vol. 16, no. 17, p. 12460, 2008.
[62] Z. Zhang, C. Yuan, S. Shen, H. Yu, R. Zhang, H. Wang, H. Li, Y. Wang, G. Deng, Z. Wang, L. You, Z. Wang, H. Song, G. Guo, and Q. Zhou, "High-performance quantum entanglement generation via cascaded second-order nonlinear processes," *npj Quantum Inf.*, vol. 7, no. 1, p. 123, Aug. 2021.
[63] O. Lib, G. Hasson, and Y. Bromberg, "Real-time shaping of entangled photons by classical control and feedback," *Sci. Adv.*, vol. 6, no. 37, Sep. 2020, Art. no. eabb6298.
[64] J. Fulconis, O. Alibart, J. L. O'Brien, W. J. Wadsworth, and J. G. Rarity, "Nonclassical interference and entanglement generation using a photonic crystal fiber pair photon source," *Phys. Rev. Lett.*, vol. 99, no. 12, Sep. 2007, Art. no. 120501.
[65] B. Lasers. (2023). *200 mw–500 mw 405 nm Violet Diode Laser System*. [Online]. Available: https://www.berlinlasers.com/200mw-500mw-405nm-violet-diode-laser-system
[66] *1.7 Watt (1700 mw) Diode Violet 405 nm Laser Module*. Endurance Lasers. Accessed: Nov. 29, 2023. [Online]. Available: https://www.endurance-lasers.com/products/1-7-watt-1700-mw-diode-violet-405-nm-laser-module
[67] LaserTack. (Nov. 2023). *448 nm Violet Laser Diode, 3500 mw to-Can Package*. Accessed: Nov. 29, 2023. [Online]. Available: https://www.laserdiodesource.com/shop/448nm-3500mw-TO-Can-Laser-Diode-Lt
[68] (2023). *Single Photon Counting Module*. Accessed: Nov. 29, 2023. [Online]. Available: https://qubitekk.com/products/single-photon-counting-module/
[69] Qubitekk. (2023). *Single Photon Counting Module*. [Online]. Available: https://qubitekk.com/products/single-photon-counting-module/
[70] (2023). *How Much Power Do Servers Consume?* Googleusercontent. [Online]. Available: https://medium.com/@zodhyatech/how-much-power-do-the-servers-consume-8f9620be71a

YASSIR AL-KARAWI (Member, IEEE) received the bachelor's degree in electronics and communications engineering from the University of Technology, Iraq, in 2002, and the M.Sc. degree in communications engineering from the University of Technology Malaysia (UTM), in 2012. He is currently pursuing the Ph.D. degree with Brunel University London, U.K., from 2020. He was an Engineer with Diyala University, Iraq, from 2006 to 2010. He has published several journals and conference papers. Also, he is a member of Staff with the Department of Communications, College of Engineering, University of Diyala. His research interests include radio frequency spectrum sharing optimizations, coexistence, and compatibility studies between cellular and fixed services, including next-generation systems and networks.

HAMED AL-RAWESHIDY (Senior Member, IEEE) received the B.Eng. and M.Sc. degrees from the University of Technology, Baghdad, Iraq, in 1977 and 1980, respectively, the master's Diploma degree from Glasgow University, Glasgow, U.K., in 1987, and the Ph.D. degree from Strathclyde University, Glasgow, in 1991. He was with the Space and Astronomy Research Centre, Baghdad; PerkinElmer, Waltham, MA, USA; British Telecom, London, U.K.; Oxford University, Oxford, U.K.; Manchester Metropolitan University, Manchester, U.K.; and Kent University, Canterbury, U.K. He is currently the Director of the Wireless Network Communications Centre, Brunel University London.

RAJAGOPAL NILAVALAN (Senior Member, IEEE) received the B.Sc. (Eng.) degree in electrical and electronics engineering from the University of Peradeniya, Sri Lanka, in 1995, and the Ph.D. degree in radio frequency systems from the University of Bristol, Bristol, U.K., in 2001. From 1999 to 2005, he was a Researcher with the Centre for Communications Research (CCR), University of Bristol, U.K. At Bristol, his research involved theoretical and practical analyses of postreception synthetic focusing concepts for near-field imaging and research on numerical FDTD techniques. Since 2005, he has been with the Department of Electronics and Computer Engineering, Brunel University London, U.K., where he is currently a Reader in wireless communications. His main research interests include antennas and propagation, microwave circuit designs, numerical electromagnetic modeling, and wireless communication systems. He has published over 100 papers and articles in journals and international conferences in his research areas. He was a member of the European commission, Network of Excellence on Antennas, from 2002 to 2005, and a member of the IET.