Journal of Engineering Science and Technology
Vol. 19, No. 1 (2024) 335 - 350

EVOLUTIONARY NETWORK SLICE ASSOCIATION ALGORITHM FOR LOAD BALANCING IN HETEROGENEOUS OPEN RADIO ACCESS NETWORKS

UWAIS AL-QARNI ANWAR RIDWAN[1],
TEONG CHEE CHUAH[1,*], YING LOONG LEE[2]

[1]Faculty of Engineering, Multimedia University, Cyberjaya, 63100, Malaysia
[2]Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Bandar Sungai Long, 43200, Malaysia
*Corresponding Author: tcchuah@mmu.edu.my

Abstract

The problem of user association (UA) for a network slicing-enabled heterogeneous network deployed under an Open Radio Access Network (O-RAN) architecture is investigated in this paper. Efficient UA is crucial for achieving load balance among network slices (NSs) and preventing overloading or underutilisation. UA is challenging in such networks because user equipment (UE) establishes wireless connectivity through NSs virtualised at the open radio units, creating a three-level association relationship. The UA problem is first formulated to maximise a generalised fairness utility function subject to constraints on NS resources and UE quality of service. To solve this computationally hard integer-programming problem, the constraints are converted into penalty functions that transform the constrained problem into an unconstrained one. A genetic algorithm (GA) is then designed to solve the unconstrained problem. Simulation results demonstrate that the proposed scheme outperforms baseline schemes in terms of load balance and quality of service fulfilment. Our work highlights the critical importance of efficient UA in heterogeneous networks with network slicing deployed under the O-RAN architecture and provides a promising framework for future research in this area.

Keywords: Genetic algorithm, Heterogeneous networks, Network slicing, O-RAN, User association.

1. Introduction

The proliferation of mobile devices and the diverse range of use cases have led to the development of 5G networks. However, the deployment of 5G networks requires significant investment in infrastructure and equipment, which can lead to vendor lock-in, high costs, and a lack of interoperability between different vendors' equipment [1]. The Open Radio Access Network (O-RAN) architecture addresses these challenges, allowing for vendor flexibility and competition. O-RANs also enable the functional split into open radio units (O-RUs), open distributed units (O-DUs), and open centralised units (O-CUs), enabling virtual network functions to be distributed across them for enhanced performance [2].

A key technique to support diverse use cases is network slicing, which enables multiple virtual networks to be created on a single physical network infrastructure, with each network slice (NS) designed to meet specific quality of service (QoS) requirements [3, 4]. However, deploying heterogeneous networks that include cells of different sizes and transmission power presents significant challenges in ensuring efficient user association, which is essential for maintaining QoS and load balance. To address this issue, user association (UA) schemes are designed to associate user equipment (UE) with the best base stations (BSs) in such a way that the loads among BSs are balanced and the QoS-requirements of UEs are satisfied.

Kim et al. [5] proposed a theoretical framework addressing the UA problem in wireless networks. Their framework specifically targeted distributed load balancing in the context of spatially inhomogeneous traffic distributions. Ye et al. [6] proposed a UA scheme that achieves load balance in heterogeneous networks by solving a network-wide utility maximisation problem. Bayat et al. [7] proposed a distributed algorithm to identify optimal UE association and femto access point allocations. Zhang et al. [8] proposed a UA algorithm that balances energy consumption and traffic loads among heterogeneous BSs relying on renewable energy supply. Hirata et al. [9] proposed a greedy heuristic method to balance the load between macro cells and small cells. Liu et al. [10] and Ramazanali et al. [11] have presented an extensive review of state-of-the-art UA algorithms for heterogeneous networks, massive multiple-input multiple-output (MIMO), millimetre wave, and energy harvesting networks. Lai et al. [12] proposed a heuristic algorithm for joint cell selection and resource allocation scheme for LTE-Advanced heterogeneous networks, while Zhao et al. [13] and Zhang et al. [14] proposed a deep learning-based UA scheme for heterogeneous networks.

Recently, some authors have also investigated UA schemes for network-slicing enabled heterogeneous networks. Amine et al. [15] proposed a new network slicing architecture to facilitate UA in 5G ultra-dense heterogeneous networks. Ye et al. [16] proposed a joint UA and resource allocation scheme for load balancing in homogeneous networks with slicing capability. Jayanthi et al. [17] proposed an evolutionary approach for UA in a multi-tenant sliced heterogeneous network. However, they assumed that each NS is served only by one BS, which is not realistic, and co-channel interference has not been taken into account. Joda et al. [18] proposed a deep reinforcement learning-based joint UA and centralised unit (CU)-distributed unit (DU) placement in O-RAN architecture by obtaining optimal placement of CU and DU network functions while jointly associating the users with radio units. However, they did not consider network slicing-enabled heterogeneous networks. Nizam et al. [19] proposed a UA scheme for network slicing enabled-hybrid wireless-wireline access networks (HWWANs). However, they did not consider the O-RAN architecture in their system model.

To date, there is a lack of research on UA schemes for network slicing-enabled heterogeneous networks deployed under an O-RAN architecture. The UA problem in such networks is further exacerbated because UE establishes wireless connectivity through NSs that have been virtualised at the O-RUs, creating a three-level association relationship. Efficient UA is crucial for achieving NS load balance and preventing overloading or underutilisation. More importantly, none of the above-mentioned UA schemes can be applied to such networks due to their inability to perform three-level associations between users, NSs, and O-RUs.

In the current paper, we aim to propose a UA scheme for such networks with the aim of achieving fair load balance among NSs while satisfying the QoS requirements of the UEs. In particular, inspired by the ability of the genetic algorithm (GA) to handle complex and nonlinear optimisation problems, we develop a QoS-aware UA scheme based on the GA, which takes into account the limited user capacity of each NS and the spectral efficiency requirements of the UEs. The contributions of this paper are: 1) The problem of load balancing in a heterogeneous O-RAN enabled with network slicing is formulated as an optimisation problem that involves three-level associations. The objective is to maximise fairness among the NSs while satisfying constraints such as limited user capacity of each NS and the minimum spectral efficiency requirements of each UE. 2) A GA-based UA algorithm is developed for solving the load balancing optimisation problem. In addition, the algorithm's complexity is analysed.

The rest of this paper is organised as follows. In the next section, we present a detailed explanation of our system model, which includes the problem formulation and the proposed UA scheme. Next, we delve into numerical simulations and performance analysis to provide a comprehensive evaluation of our proposed approach. In the last section, we present our conclusion, which summarises the key findings of our research and offers insights into potential areas for future exploration.

2. Methods

This section presents the system model, problem formulation and proposed genetic algorithm for user association.

2.1. System model and problem formulation

The network considered in this study is a sliced heterogeneous network comprising a macro O-RU and multiple small-cell O-RUs, as depicted in Fig. 1. Each NS is managed by a mobile virtual network operator (MVNO) and served by multiple BSs sharing their radio resource blocks (RBs). We assume the spectrum is split between the macro BS (MBS) and the small BSs (SBSs), resulting in a frequency reuse of half. Due to the deployment strategy of the respective MVNO, not all NSs provisioned through every O-RU. For example, NS 1 is accessible via the macro O-RU 1 but not the small-cell O-RU 2, while NS 2 is accessible via both O-RU 1 and O-RU 2. The sets of O-RUs, UE, and NSs are denoted by $K = \{1, 2, ..., k, ..., |K|\}$, $U = \{1, 2, ..., u, ..., |U|\}$, and $M = \{1, 2, ..., m, ..., |M|\}$, respectively, and the association between each NS and each O-RU is described by

$$x_{km} = \begin{cases} 1 & \text{if NS } m \text{ is accessible via BS } k \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

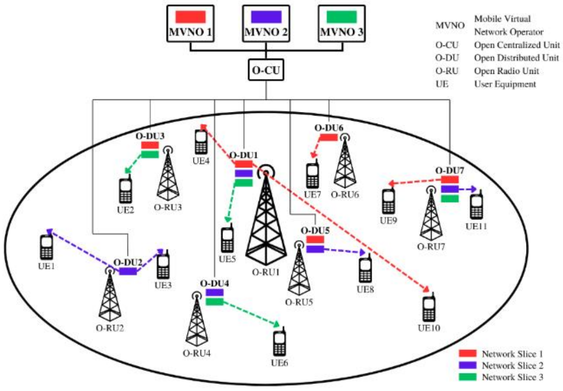

Fig. 1. System model of a network slicing-enabled heterogeneous open radio access network.

It is assumed that the association between O-RUs and NSs has been pre-negotiated between the MVNOs and the infrastructural network provider, that is, the values of $\{x_{km}\}$ are known a priori. Each NS m is identified by its processing capacity that can be provided to each of its associated UE in bits/second [20], denoted as R_m, and the number of UEs that can be supported by NS m is denoted as V_m. Next, we define the UA variable as

$$a_{kmu} = \begin{cases} 1 & \text{if user } u \text{ associates with NS } m \text{ via BS } k \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

By taking into account the interference between O-RUs and assuming equal distribution of transmit power across all bandwidth in the associated NS, the achievable data rate of user $u \in U$ over the system bandwidth if it associates with O-RU $k \in K$ can be modelled as

$$R_{ku} = B\log_2\left(1 + \frac{P_{\max,k}\, G_{ku}}{\sum_{i\in K\{k\}} P_{\max,i} G_{iu} + N_0}\right)\text{bits/s} \tag{3}$$

where B is the system bandwidth; $P_{\max,k}$ is the maximum transmit power of BS k; G_{ku} is the channel gain between O-RU k and UE u, which is averaged over the UE association interval; N_0 is the additive white Gaussian noise power; and $\sum_{i\in K\{k\}} P_{\max,i} G_{iu}$ is the worst-case interference experienced by UE u if it associates with O-RU k.

Next, we define the load of each NS as the ratio of the number of UEs associated with the NS to its maximum user capacity, that is

$$\eta_m = \frac{\sum_{k\in K}\sum_{u\in U} x_{mk} a_{kmu}}{V_m} \tag{4}$$

where $\sum_{k\in K}\sum_{u\in U} x_{mk} a_{kmu}$ is the number of UEs associated with NS m. In the current study, we aim to balance the loads among all the NSs in order to achieve a fair load distribution among NSs and efficient resource utilisation. To this end, we formulate the load balancing problem as a mathematical problem that maximises a generalised fairness utility function, namely α-fairness utility function with respect to the number of UE associated with the NSs in the sliced heterogeneous network. The maximisation can result in α-fair UA among the NSs and is expressed as follows:

$$\max_{\{a_{kmu}\}} \sum_{m\in M} f_\alpha(\eta_m), \tag{5}$$

subject to

$$\sum_{m\in M}\sum_{k\in K} x_{km}a_{kmu}R_m \geq R_{\min,u} \ \ \forall u \in U \tag{5a}$$

$$\sum_{m\in M}\sum_{k\in K} x_{km}a_{kmu}R_{ku} \geq R_{\min,u} \ \ \forall u \in U \tag{5b}$$

$$\sum_{k\in K}\sum_{u\in U} x_{mk}a_{kmu} \leq V_m \ \ \forall m \in M \tag{5c}$$

$$\sum_{m\in M}\sum_{k\in K} x_{km}a_{kmu} \leq 1 \ \ \forall u \in U \tag{5d}$$

$$a_{kmu} \in \{0,1\} \ \ \forall u \in U, k \in K, m \in M \tag{5e}$$

where

$$f_\alpha(\eta_m) = \begin{cases} \frac{{\eta_m}^{1-\alpha}}{1-\alpha} & \alpha \neq 1 \\ \log(\eta_m) & \alpha = 1 \end{cases} \tag{6}$$

The fairness notion considered is indicated by α. For instance, proportional fairness corresponds to $\alpha = 1$, delay-fairness to $\alpha = 2$, and max-min fairness to $\alpha \rightarrow \infty$ [21]. Constraint Eqs. (5a) and (5b) ensure that each UE u is associated with a NS m via O-RU k that can satisfy its minimum spectral efficiency requirement $R_{\min,u}$. Constraint Eq. (5c) ensures that the number of UE associated with NS m does not exceed its maximum user capacity V_m. Constraint Eq. (5d) ensures that each UE only associates with one NS through one O-RU at one time. Constraint Eq. (5e) specifies that a_{kmu} is a binary variable that can only take the values of zero or one.

2.2. Proposed genetic algorithm for user association

It is observed that the optimisation problem in Eq. (5) is a 0-1 integer programming problem, which is computationally NP-hard [22]. To solve this problem, we leverage the GA, which draws inspiration from biological evolution and natural selection, and it has been shown to be effective in solving unconstrained large-scale 0-1 integer programming problems. However, Eq. (5) presents a complex optimisation problem with multiple constraints, making it unsuitable for direct optimisation using a GA. To address this issue, we employ a transformation technique to convert the constrained optimisation problem in Eq. (5) into an unconstrained one. To this end, we leverage the penalty function approach to convert constraints Eqs. (5a) - (5d) into penalty functions. Firstly, we convert Eqs. (5a) and (5b) as follows:

$$\sum_{m\in M}\sum_{k\in K} x_{km}a_{kmu}\frac{R_m}{B} \geq \frac{R_{\min,u}}{B} \ \forall u \in U \tag{7a}$$

$$\sum_{m\in M}\sum_{k\in K} x_{km}a_{kmu}S_{ku} \geq \frac{R_{\min,u}}{B} \ \ \forall u \in U \tag{7b}$$

where $S_{ku} = \log_2\left(1 + \frac{P_{\max,k}\, G_{ku}}{\sum_{i\in K\{k\}} P_{\max,i}G_{iu}+N_0}\right)$. Then, we convert Eqs. (5c), (5d), (7a) and (7b) into a single penalty function:

$$p(a_{kmu}) = \langle p_1(a_{kmu})\rangle + \langle p_2(a_{kmu})\rangle + \langle p_3(a_{kmu})\rangle + \langle p_4(a_{kmu})\rangle \tag{8}$$

where

$$p_1(a_{kmu}) = w_1\left(\sum_{u\in U}\left(\sum_{m\in M}\sum_{k\in K} x_{km}a_{kmu}\frac{R_m}{B}\right) - \frac{R_{\min,u}}{B}\right), \tag{8a}$$

$$p_2(a_{kmu}) = w_2\left(\sum_{u\in U}\left(\sum_{m\in M}\sum_{k\in K} a_{kmu}x_{km}S_{ku}\right) - \frac{R_{\min,u}}{B}\right), \tag{8b}$$

$$p_3(a_{kmu}) = w_3\left(\sum_{m\in M}\left(V_m - \sum_{k\in K}\sum_{u\in U} x_{km}a_{kmu}\right)\right), \tag{8c}$$

$$p_4(a_{kmu}) = w_4(\sum_{u\in U}(1 - \sum_{k\in K}\sum_{u\in U} x_{km}a_{kmu})\ ), \tag{8d}$$

with

$$\langle p_i(a_{kmu})\rangle = \begin{cases} p_i(a_{kmu}) & p_i(a_{kmu}) < 0 \\ 0 & p_i(a_{kmu}) \geq 0 \end{cases}, i = 1, 2, 3, 4 \tag{9}$$

where w_1, w_2, w_3 and w_4 are the penalty constants. Equations (8a), (8b), (8c) and (8d) correspond to the constraints in Eqs. (5a), (5b), (5c) and (5d) respectively with an additional piecewise constraint expressed in Eq. (9). Then, we transform the problem in Eq. (5) by incorporating the penalty functions in Eq. (8) into the objective function of Eq. (5) [23], forming the following unconstrained problem:

$$\max_{\{a_{kmu}\}}\phi(a_{kmu}) = f(a_{kmu}) + p(a_{kmu}) \tag{10}$$

where

$$f(a_{kmu}) = \sum_{m\in M} f_\alpha\left(\frac{\sum_{k\in K}\sum_{u\in U} x_{mk}a_{kmu}}{V_m}\right) \tag{11}$$

It is worth noting that when any of the constraints in Eqs. (5a) - (5d) are violated, the value of $p(a_{kmu})$ will be a very large negative value, which will act as a 'penalise' term, limiting the maximisation of the problem.

Next, with the unconstrained problem in Eq. (10), we can apply the GA to solve the optimisation problem. The GA optimisation process starts with generating an initial population of candidate solutions. This population is composed of the UA matrix, a_{kmu}, which is defined in Eq. (2). This matrix can be perceived as a 3-dimensional array with dimensions [$k \times m \times u$]. In order to utilise the GA, it is necessary to convert this 3-dimensional array UA matrix a_{kmu} into a 1-dimensional array $v_{a_{kmu}}$, also known as 'chromosomes', through as a vectorisation process outlined in Eq. (12). The overall objective function can then be incorporated into Eq. (13).

$$v_{a_{kmu}} = vectorization\ (a_{kmu}) \tag{12}$$

$$\phi(v_{a_{kmu}}) = f(v_{a_{kmu}}) + p(v_{a_{kmu}}) \tag{13}$$

Once the fitness of each individual has been evaluated using the objective function, a selection process is employed to choose parents for reproduction. This process utilises the roulette wheel selection method, which considers the fitness of each individual in the population as a probability for selection. The selection pressure, β, is used to balance the exploration and exploitation in the search process. In the GA, new individuals are created by applying genetic operators such as crossover and mutation, and their fitness is evaluated using the objective function.

We consider a hybrid approach of crossover which combines single-point, double-point, and uniform crossover. The probability of selected parents exchanging genetic information is controlled by a crossover probability (p_C), while the probability of a gene in a chromosome being randomly altered is determined by a mutation probability (μ). Additionally, to further enhance the diversity of the population, new individuals are created using different random generators from random number generators [24], with the number of such individuals denoted by n_P. However, all new individuals must comply with the constraints for optimal results.

Additionally, a replacement strategy is used to substitute the old population with newly generated individuals obtained through genetic operators. The newly created individuals are then assessed for their fitness. This iterative process continues until a stopping criterion is satisfied, such as reaching the maximum number of iterations denoted by T_{Max}. In summary, the implementation of a GA comprises the following steps: defining the optimisation problem, initialising the population, evaluating fitness, selecting parents, reproducing new individuals via reproduction and mutation, introducing random individuals, evaluating their fitness, replacing the old population, and iterating the process until the stopping criterion is satisfied. Table 1 describes the flow of the proposed GA-based UA scheme.

Table 1. Proposed GA-based UA scheme.

Step 1: Set the GA parameters such as the number of individuals in a population (n_{Pop}), maximum number of iterations (T_{Max}), probability of crossover (p_C), probability of mutation (μ), pressure of selection (β), number of new individuals (n_P) and penalty constants ($w_{1,}$ $w_{2,}$ w_3 and w_4).
Step 2: Randomly generate a population of UA matrix with a dimension of [$k \times m \times u$]. Vectorise each UA matrix to form a chromosome as in Eq. (11).
Step 3: Compute the fitness score using Eq. (12), and then rank the chromosomes based on their fitness scores.
Step 4: Generate new individuals with the same dimension as in **Step 2** and use the Roulette Selection Wheel method to choose an even number of chromosomes with high fitness values.
Step 5: Perform GA operations of crossover and mutation.
Step 6: Evaluate the fitness score of each candidate solution. Then sort the population and trim it by removing the least fit individuals.
Step 7: If the iteration number is equal to the predefined maximum number of iterations T_{Max}, terminate the algorithm; otherwise go to **Step 4** with the new candidate solutions.

The proposed GA-based UA scheme can be implemented in a centralised manner at the O-CU as shown in Fig. 2, which is responsible for managing and controlling the radio access network (RAN) functions. In particular, the O-CU provides a centralised control plane that manages and coordinates all the network functions and resources, including radio resource management, mobility management, and security management [25]. As a result, it plays a critical role in optimising network performance by providing the central intelligence and decision-making capability necessary to ensure efficient network operation. The computational complexity of the proposed GA-based UA scheme is $O(T_{\mathrm{Max}} n_{\mathrm{Pop}} |K||M||U|)$, which is reasonable and can be handled efficiently by modern computing resources.

The overall flowchart of the proposed GA-based UA scheme is shown in Fig. 2. Initially, the O-CU analyses the network environment, including the number of BSs, NSs, and UEs, as given in Table 2. The O-CU then communicates with the O-DUs and O-RUs to collect the channel gain information for individual UE. Next, the proposed UA scheme is implemented using the network state to optimise UA based on the predefined criteria. To begin, a population of UA matrices is generated, each of which is vectorised to create a one-dimensional chromosome, and its fitness is computed. The algorithm then uses a Roulette Wheel Selection method to select high-fitness matrices as parents for the hybrid crossover function, which includes single and double-point crossover and uniform crossover. Occasionally, mutation is applied to increase the genetic variability of the population. The fitness of the new individuals is evaluated, and the population is trimmed for a constant number of individuals in each generation. This iterative process is repeated until the maximum number of generations is achieved, and the O-CU produces an optimised UA matrix that can adapt to the changing network conditions. With the GA-based UA scheme, the O-CU ensures efficient utilisation of network resources and provides high-quality service to users.

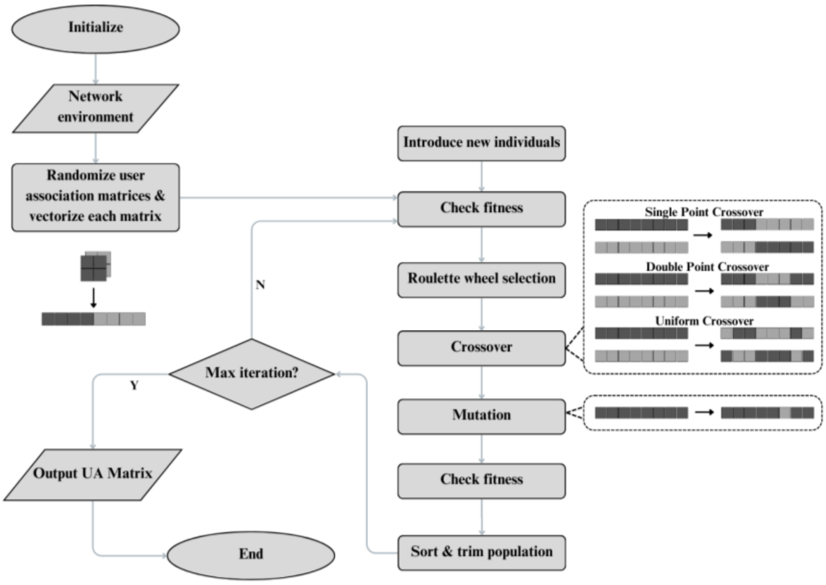

Fig. 2. Flowchart of the proposed GA-based UA scheme.

3. Results and Discussion

The performance of the proposed GA-based UA scheme is evaluated for a heterogeneous O-RAN that utilises a bandwidth of B = 20 MHz which is split into two spectra of 10 MHz each, with one for the macro O-RU and the other for the 19 small-cell O-RUs. All the small-cell O-RUs are deployed within the coverage area of the macro-cell O-RU, which has a radius of 500 m. We assume the deployment of three NSs or MVNOs, with R_1 = 15 Mbits/s, R_2 = 20 Mbits/s, and R_3 = 25 Mbits/s, respectively. The assumed bit rates are aligned with enhanced mobile broadband (eMBB) data rate requirements, specifically addressing live video streaming bit rate needs. We refer to Table 3 in reference [26] for the ranges of the bit rate requirements. The maximum user capacity of each NS, V_m, is set to 50. We randomly assign a unique combination of NSs to each O-DU, such that some O-DUs may have NSs 1 and 2, while others may have NSs 1 and 3, and so on.

We assume that the macro-cell O-RU and small-cell O-RUs transmit signals at power levels of 43 dBm and 30 dBm, respectively, by following reference [20]. We use distinct path loss models for macro and small cells. Adopting 3GPP-standardized path loss models from [27], the path loss for the macro cell is 140.7 + 36.7 log(d) dB, while for small cells, it is 128.1 + 37.6 log(d) dB, where d denotes the distance in kilometers between the O-RU and the UEs. The total number of UEs is systematically varied at values of 50, 100, 150, and 200. It is assumed that the UEs are randomly distributed within the coverage area of the macro-cell O-RU. Additionally, we consider a communication channel with zero-mean unit-variance Rayleigh fading, and zero-mean log-normal shadowing with a 10 dB standard deviation. The noise power spectral density is set to -174 dBm/Hz, while the antenna gain, and noise figure are set to 5 dB and 9 dB, respectively. The minimum user spectral efficiency requirement $R_{\min}$ is set randomly between 1 bit/s/Hz to 2.5 bits/s/Hz. The alpha value α for the fairness utility function is set to 1 for proportional fairness. The simulation parameters used are summarised in Table 2.

Table 2. Simulation parameters used to assess the performance of the proposed GA-based UA scheme.

Parameters	Value
Macro cell radius	500 m
System bandwidth	20 MHz (10 MHz - Macro, 10 MHz - Small)
Antenna gain	5 dB
MBS transmit power	43 dBm
SBS transmit power	25 dBm
Noise power density	-174 dBm/Hz
Noise figure	9 dB
MBS pathloss model	140.7 + 36.7 log(d) dB
SBS pathloss model	128.1 + 37.6 log(d) dB
Shadowing standard deviation	10 dB
Number of BSs	20 (1 - Macro, 19 - Small)
Number of MVNOs	3
Number of NSs	3
Number of UEs	50 - 200
Guaranteed data rate by NSs, R_m	15 Mbits/s, 20 Mbits/s, 25 Mbits/s
Maximum user capacity of each NS, V_m	50
Minimum user spectral efficiency requirement, $R_{\min}$	1 bit/s/Hz - 2.5 bits/s/Hz
Alpha value, α	1
Macro cell radius	500 m

For the proposed GA scheme, the settings for the GA are user-defined and are determined experimentally through extensive testing of various configurations. We set the maximum number of iterations T_{Max} to 500, and the population size n_{Pop} to 20 chromosomes. Each chromosome represents a randomised UA matrix that has been converted. To ensure that each matrix satisfies the association constraint linking the NS with the O-RU, we randomise the UA matrices while ensuring compliance with this constraint. The GA operates with a crossover probability p_C of 1, a mutation probability μ of 0.02, and a selection pressure β of 1. In order to promote greater genetic diversity within the population, we propose the addition of two new individuals to the GA. By introducing these new individuals, denoted by $n_P = 2$, we aim to increase the genetic variability of the population and facilitate more thorough exploration of the solution space. This can help to prevent premature convergence and improve the algorithm's ability to identify high-quality solutions. Additionally, we prioritise the cost function over other penalty functions during the calculation by setting the penalty constants w_1, w_2, w_3 and w_4 to 0.001. The parameters of the GA are listed in Table 3.

The performance of the proposed GA is first evaluated by analysing its convergence behaviour for different total numbers of UEs, as shown in Fig. 3. The results demonstrate the effectiveness of the proposed GA in solving the UA optimisation problem, as evidenced by the improvement of the cost function with each iteration. Additionally, the best cost value increases with the number of UEs, underscoring the significance of fairness and connectivity. The cost function is more significant when a higher the number of UEs can connect and meet the maximum allowable connections of the NS. Nevertheless, as the number of UEs increases, the improvement in the best cost values becomes less significant. This suggests that the network capacity constraints for larger UEs numbers are more stringent, with the network resources being adequate for only approximately 200 UEs.

Table 3. GA parameters.

Parameters	Value
Maximum number of iterations, T_{Max}	500
Population size, n_{Pop}	20
Crossover probability, p_C	1
Mutation probability, μ	0.02
Selection pressure, β	1
Number of new individuals, n_P	2
Penalty constants, w	0.001

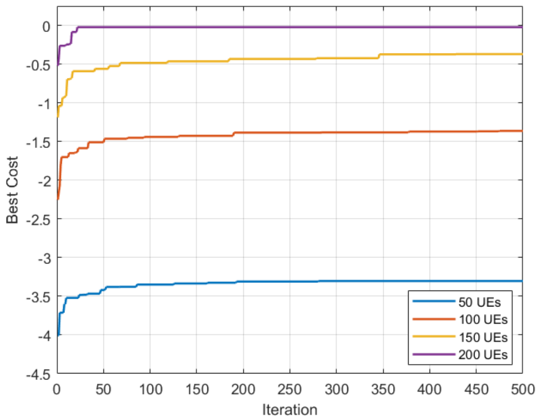

Fig. 3. Convergence curves of the proposed GA-based UA scheme.

To benchmark the performance of our proposed scheme, we compare it with two baseline schemes. The first baseline scheme, denoted as BSNS, associates the UE with the O-RUs based on the maximum signal-to-interference-plus-noise ratio (SINR) and then assigns the UE to a NS available in the associated O-RU. The second baseline scheme, denoted as NSBS, assigns the UEs to a NS that can fulfil the target data rates, and then associates it with an O-RU where the associated NS is accessible based on the maximum SINR. To ensure the reliability of our results, we conducted 100 simulations and computed the average of the obtained results.

Figure 4 shows the percentage of QoS-satisfied UEs, namely UEs whose minimum spectral efficiency requirements are met. It can be calculated as $\gamma = \frac{\sum_{m\in M}\sum_{k\in K}\sum_{u\in U} x_{mk} a_{kmu}}{U} \times 100\%$, subject to the constraints in Eqs. (5a) to (5e). Figure 4 demonstrates that the proposed scheme outperforms both the BSNS and NSBS schemes by almost 15% to 30%.

This can be attributed to the algorithm's capability to explore a larger search space, resulting in better solutions compared to the baseline counterparts. Notably, the proposed scheme exhibits exceptional performance due to its consideration of various factors during the UA process, including the spectral efficiency requirements of UEs and the user capacity of NSs. The proposed scheme results in an increased number of UEs meeting their QoS requirements and leads to improved load balancing, making it a highly effective solution in comparison to the baseline schemes.

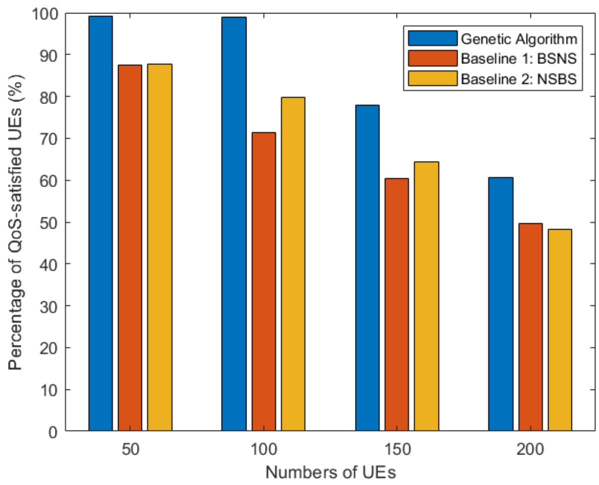

Fig. 4. Performance comparison in terms of QoS-satisfied UEs.

To assess the load balance of the network, Fig. 5 compares the performance of the three schemes using Jain's fairness index (FI) [28], which is defined as

$$\mathrm{FI} = \frac{(\sum_{m\epsilon M}\eta_m)^2}{|M|\sum_{m\epsilon M}(\eta_m)^2} \tag{14}$$

The results indicate that the proposed scheme outperforms both baseline schemes in terms of the FI. This demonstrates the effectiveness of the proposed scheme in achieving load balancing by maximising the fairness utility function in Eq. (5), leading to a more equitable distribution of services among the NSs and contributing to a better-balanced network.

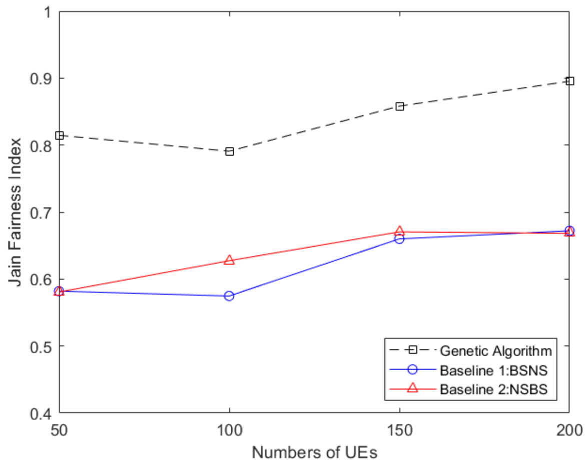

Fig. 5. Performance comparison in terms of Jain's fairness index.

Finally, we investigate the performance of the three schemes in terms of average spectral efficiency and sum spectral efficiency, which can be calculated respectively, as follows:

$$\eta_{\mathrm{avg}} = \frac{\sum_{m\in M}\sum_{k\in K}\sum_{u\in U} x_{km}a_{kmu}R}{U} \tag{15}$$

$$\eta_{sum} = \sum_{m \in M} \sum_{k \in K} \sum_{u \in U} x_{km} a_{kmu} R \quad (16)$$

where

$$R = \begin{cases} \frac{R_m}{B} & , \frac{R_m}{B} < S_{ku} \\ S_{ku} & , \frac{R_m}{B} > S_{ku} \end{cases} \quad (17)$$

Figure 6 shows performance of the three schemes in terms of average spectral efficiency. The results show that the proposed algorithm, which prioritises proportional fairness and minimises the number of dropped UE, maintains a high average spectral efficiency for the UEs while successfully achieving its objective. The proposed algorithm outperforms the baseline schemes by achieving a slightly higher average spectral efficiency.

Figure 7 compares the sum spectral efficiency performance among the three schemes, representing the aggregate spectral efficiency of all UEs. The comparison reveals that the proposed algorithm outperforms the baseline schemes. In particular, the baseline schemes have been found to underperform and are inefficient compared to the proposed UA algorithm. This is because the baseline schemes focus on maximising signal strength and do not prioritise load balancing, leading to an uneven distribution of traffic and the possibility of congestion in specific areas of the network.

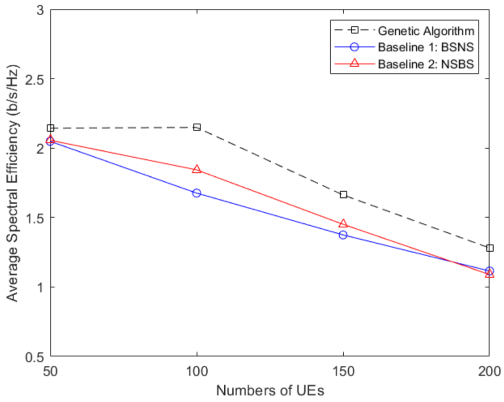

Fig. 6. Performance comparison in terms of average spectral efficiency.

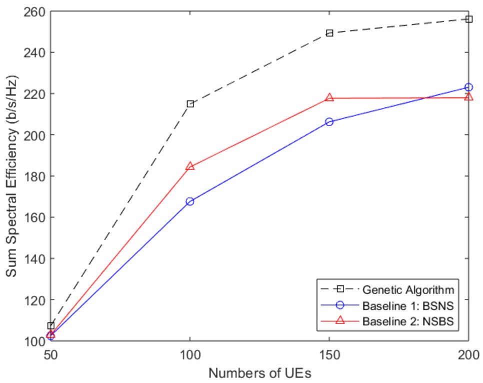

Fig. 7. Performance comparison in terms of sum spectral efficiency.

Additionally, the baseline schemes do not take into account the varying needs different user types, resulting in an inability to provide adequate QoS for all users. In contrast, the proposed UA algorithm takes into account both load balancing and QoS requirements, leading to a more even distribution of traffic and overall improvement in network performance. By using the proposed algorithm, network operators can achieve better resource allocation and management, leading to a more efficient and effective network that meets the needs of all users. This algorithm prioritises the optimisation of network performance and the enhancement of user experience by considering both QoS requirements and the load balancing of BSs. By leveraging network slicing capabilities, the proposed algorithm facilitates a more efficient utilisation of resources and a better allocation of traffic, resulting in a significant increase in network capacity and user satisfaction overall.

4. Conclusion

In this paper, we addressed the critical issue of UA for a network slicing-enabled heterogeneous network deployed under the O-RAN architecture. The three-level association relationship between the UEs, NSs, and O-RUs in network slicing-enabled heterogeneous networks makes UA challenging. To solve this complex problem, we formulated the UA problem to achieve proportional-fair load balancing among NSs and solved it using a GA. Simulation result demonstrated that our proposed scheme outperforms two baseline approaches in terms of fairness and QoS-fulfilment as well as maintains high average spectral efficiency provided to the UEs. Future work could consider further optimisation of the proposed UA approach for improved energy-efficiency as well as its extension to more complex network scenarios.

Acknowledgement

This work was supported in part by the Ministry of Higher Education Malaysia under Grant FRGS/1/2019/ICT05/MMU/01/1 and Multimedia University.

Nomenclatures

a_{kmu}	UA variable
B	System bandwidth, Hz
FI	Fairness index
G_{ku}	Channel gain between O-RU *k* and UE *u*, dB
k	Index number of the BS
m	Index number of the NS
N_0	Additive white Gaussian noise power
n_P	Number of new individuals
n_{Pop}	Population size used by the GA
$O()$	Big O notation
p_C	Crossover probability used by the GA
$P_{\max,k}$	Maximum transmit power of BS *k*, dBm
R_{ku}	Achievable data rate of UE *u* when it associates with O-RU *k*, b/s
R_m	Data rate guaranteed by NS *m*, Mb/s
$R_{\min}$	Minimum user spectral efficiency requirement, b/s/Hz
S_{ku}	Achievable spectral efficiency of UE *u* if it associates with O-RU *k*, b/s/Hz
T_{Max}	Maximum number of iterations of the GA in each generation

u	Index number of the UE
V_m	Maximum user capacity of each NS
x_{km}	Association variable between NS *m* and O-RU *k*

Greek Symbols

α	Alpha value (type of fairness)
β	Selection pressure
μ	Mutation probability
γ	Percentage of QoS-satisfied UEs
η_{avg}	Average spectral efficiency, b/s/Hz
η_m	Load of NS *m* as the ratio of the number of UEs associated with NS *m* to its maximum user capacity
η_{sum}	Sum spectral efficiency, b/s/Hz

Abbreviations

BS	Base station
GA	Genetic algorithm
MBS	Macro base station
MVNO	Mobile virtual network operator
NS	Network slice
O-CU	Open centralised unit
O-DU	Open distributed unit
O-RAN	Open radio access network
O-RU	Open radio unit
SBS	Small base station
UA	User association
UE	User equipment

References

1. Shafi, M.; Molisch, A.F.; Smith, P.J.; Haustein, T.; Zhu, P.; Silva, P.D.; Tufvesson, F.; Benjebbour, A.; and Wunder, G. (2017). 5G: A tutorial overview of standards, trials, challenges, deployment, and practice. *IEEE Journal on Selected Areas in Communications*, 35(6),1201-1221.
2. Masur, P.H.; Reed, J.H.; and Tripathi, N.K. (2022). Artificial intelligence in open-radio access network. *IEEE Aerospace and Electronic Systems Magazine*, 37(9), 6-15.
3. Chartsias, P.K.; Amiras, A.; Plevrakis, I.; Samaras, I.; Katsaros, K.; Kritharidis, D.; Trouva, E.; Angelopoulos, I.; Kourtis, A.; Siddiqui, M.S.; Viñes, A.; and Escalona, E. (2017). SDN/NFV-based end to end network slicing for 5G multi-tenant networks. *Proceedings of the 2017 European Conference on Networks and Communications (EuCNC)*. Oulu, Finland, 1-5.
4. Oladejo, S.O.; and Falowo, O.E. (2017). 5G network slicing: A multi-tenancy scenario. *Proceedings of the 2017 Global Wireless Summit (GWS)*. Cape Town, South Africa, 88-92.
5. Kim, H.; de Veciana, G.; Yang, X.; and Venkatachalam, M. (2012). Distributed α-optimal user association and cell load balancing in wireless networks. *IEEE/ACM Transactions on Networking*, 20(1), 177-190.

6. Ye, Q.; Rong, B.; Chen, Y.; Al-Shalash, M.; Caramanis, C.; and Andrews, J.G. (2013). User association for load balancing in heterogeneous cellular networks. *IEEE Transactions on Wireless Communications*, 12(6), 2706-2716.
7. Bayat, S.; Louie, R.H.; Han, Z.; Vucetic, B.; and Li, Y. (2014). Distributed user association and femtocell allocation in heterogeneous wireless networks. *IEEE Transactions on Communications*, 62(8), 3027-3043.
8. Zhang, T.; Xu, H.; Liu, D.; Beaulieu, N.C.; and Zhu, Y. (2015). User association for energy-load tradeoffs in hetnets with renewable energy supply. *IEEE Communications Letters*, 19(12), 2214-2217.
9. Hirata, A.T.; Xavier, E.C.; and Borin, J.F. (2016). Load balance and user association on hetnets. *IEEE Latin America Transactions*, 14(12), 4781-4786.
10. Liu, D.; Wang, L.; Chen, Y.; Elkashlan, M.; Wong, K.-K.; Schober, R.; and Hanzo, L. (2016). User association in 5g networks: A survey and an outlook. *IEEE Communications Surveys & Tutorials*, 18(2), 1018-1044.
11. Ramazanali, H.; Mesodiakaki, A.; Vinel, A.; and Verikoukis, C. (2016). Survey of user association in 5g hetnets. *Proceedings of the 2016 8th IEEE Latin-American Conference on Communications (LATINCOM)*. Medellin, Colombia, 1-6.
12. Lai, W.K.; and Liu, J.-K. (2018). Cell selection and resource allocation in LTE-advanced heterogeneous networks. *IEEE Access*, 6, 72978-72991.
13. Zhao, N.; Liang, Y.-C.; Niyato, D.; Pei, Y.; Wu, M.; and Jiang, Y. (2019). Deep reinforcement learning for user association and resource allocation in heterogeneous cellular networks. *IEEE Transactions on Wireless Communications*, 18(11), 5141-5152.
14. Zhang, Y.; Xiong, L.; and Yu, J. (2020) Deep learning based user association in heterogeneous wireless networks. *IEEE Access*, 8, 197439-197447.
15. Amine, M.; Kobbane, A.; and Ben-Othman, J. (2020). New network slicing scheme for UE association solution in 5g ultra dense hetnets. *Proceedings of the ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*. Dublin, Ireland, 1-6.
16. Ye, Y.; Zhang, T.; and Yang, L. (2021). Joint user association and resource allocation for load balancing in ran slicing. *Physical Communication*, 49, 101459.
17. Jayanthi, S.S.; Lee, Y.L.; and Chang, Y.C. (2021). User association for multi-tenant heterogeneous network slicing using genetic algorithm. *Proceedings of the 2021 8th International Conference on Computer and Communication Engineering (ICCCE)*, Kuala Lumpur, Malaysia, 326-330.
18. Joda, R.; Pamuklu, T.; Iturria-Rivera, P.E.; and Erol-Kantarci, M. (2022). Deep reinforcement learning-based joint user association and cu-du placement in o-ran. *IEEE Transactions on Network and Service Management*, 19(4), 4097-4110.
19. Nizam, F.; Chuah, T.C.; and Lee, Y.L. (2023). User association for network slicing-enabled heterogeneous hybrid wireless-wireline access networks. *Proceedings of the 2023 International Conference on Cyber Management and Engineering (CyMaEn)*, Bangkok, Thailand, 61-65.
20. Lee, Y.L.; Loo, J.; Chuah, T.C.; and Wang, L.-C. (2018). Dynamic network slicing for multitenant heterogeneous cloud radio access networks. *IEEE Transactions on Wireless Communications*, 17(4), 2146-2161.

21. Mo, J.; and Walrand, J. (2000). Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5), 556-567.
22. Garey, M.R.; and Johnson, D.S. (1979). *Computers and intractability a guide to the theory of NP-completeness*. W.H. Freeman and Company.
23. Rao, S.S. (2009). *Engineering optimization: Theory and practice*. John Wiley and Sons.
24. Chambers, L. (2001). *The practical handbook of genetic algorithms: Applications*. Chapman & Hall/CRC.
25. Polese, M.; Bonati, L.; D'Oro, S.; Basagni, S.; and Melodia, T. (2023). Understanding o-ran: architecture, interfaces, algorithms, security, and research challenges. *IEEE Communications Surveys & Tutorials*, 25(2), 1376-1411.
26. Jiménez, L.R.; Solera, M.; Toril, M.; Luna-Ramırez, S.; and Bejarano-Luque, J.L. (2021). The upstream matters: Impact of uplink performance on YouTube 360° live video streaming in LTE. *IEEE Access*, 9, 123245-123259.
27. 3GPP (2013). Document tr 36.872, 3gpp: Small cell enhancements for E-UTRA and E-UTRAN - Physical layer aspects (Release 12). *Technical report*. Retrieved November 25, 2023, from https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2573
28. Jain, R. (1991). *The art of computer systems performance analysis*. (1st ed.). Wiley.