Alam et al. J Wireless Com Network (2024) 2024:44
https://doi.org/10.1186/s13638-024-02376-2

EURASIP Journal on Wireless
Communications and Networking

RESEARCH Open Access

# Ant colony optimization-based solution to optimize load balancing and throughput for 5G and beyond heterogeneous networks

Mohammed Jaber Alam[1*], Ritesh Chugh[2], Salahuddin Azad[2] and Md Rahat Hossain[1]

*Correspondence:
m.alam@cqu.edu.au

[1] School of Engineering and Technology, CQUniversity, North Rockhampton, QLD 4701, Australia
[2] School of Engineering and Technology, CQUniversity, Melbourne, VIC 3000, Australia

**Abstract**
The escalating demand for data in wireless communication systems has posed significant challenges in recent years. This trend is predicted to continue, with explosive data usage and evolving quality of service demands from mobile users. The rapid increase in traffic demand, combined with the intricate nature of heterogeneous network (HetNet) scenarios, has significantly heightened the challenges confronting mobile network operators. These challenges encompass service quality, load distribution, coverage, and the overall user experience. Conventional approaches that prioritize maximum received power in the cell association mechanism tend to sustain network imbalances within the HetNets, making it difficult to cater for the diverse traffic requirements of mobile users. In this study, instead of focusing solely on enhancing individual user downlink rates, we maximize the number of users whose downlink needs are satisfied by integrating a cell range extension (CRE) technique with an ant colony optimization algorithm. Our proposed method considers both the workload of base stations and the signal to interference-plus-noise ratio of user devices to formulate an objective function aimed at calculating specific CRE bias values for individual small base stations. A comparative analysis of the proposed approach with existing techniques demonstrates its effectiveness. Simulation results underscore the success of our proposed strategy in meeting users' throughput needs while reducing network imbalances and call drop rates.

**Keywords:** 5G and beyond, Heterogeneous network, Load balancing, Throughput, Ant colony optimization

## 1 Introduction

The ever-evolving landscape of wireless communication technologies has led to the emergence of 5G and beyond heterogeneous networks (HetNets), which promise extraordinary connectivity and data exchange levels. These networks are envisioned to cater for a diverse range of services, from ultra-high-speed data transmission to real-time applications and the Internet of Things (IoT) [1, 2]. However, realizing the full potential of these networks requires addressing complex challenges such as efficient user association and load balancing.

The increased demand for substantial data transmission over mobile networks has primarily been fueled by the surge in applications such as ultra-high-definition video platforms, augmented and virtual reality, and other fast internet services [3, 4]. As a result, this trend is compelling mobile network operators (MNOs) to enhance the provisioning and management of broadband services with stricter quality of service (QoS) prerequisites. While the current infrastructure of mobile broadband networks can effectively cater to several specific applications' requirements, future high-demand services will exert added demands on MNOs for the forthcoming generations of cellular networks. As such, the integration of small base stations (SBSs) with low transmission power and small coverage into macro base stations (MBS) has created ultra-dense networks (UDNs) to accommodate future network services. By strategically deploying SBSs, MNOs can expand their network coverage to the proximity of user devices, enhancing overall performance, spectral efficiency, and service deliverability [5, 6]. Figure 1 illustrates an example of a heterogeneous network consisting of a couple of SBSs overlaid inside an MBS.

The emerging need for high data traffic capacity has prompted the 3rd Generation Partnership Project (3GPP) to propose multi-tier HetNets, wherein overlaying SBSs are deployed inside conventional macro base stations (MBS) to satisfy both data rate and traffic volume requisites. This heterogeneous framework thus creates an ultra-dense network (UDN). The deployment of SBSs could address indoor coverage gaps and gaps at cell edges, providing a cost-efficient expansion of coverage [7]. Nonetheless, intensifying network density with diverse types of base stations could potentially introduce significant network imbalances, as MBSs generally operate with considerably higher transmission power than SBSs [8]. Employing the established cell association mechanism of

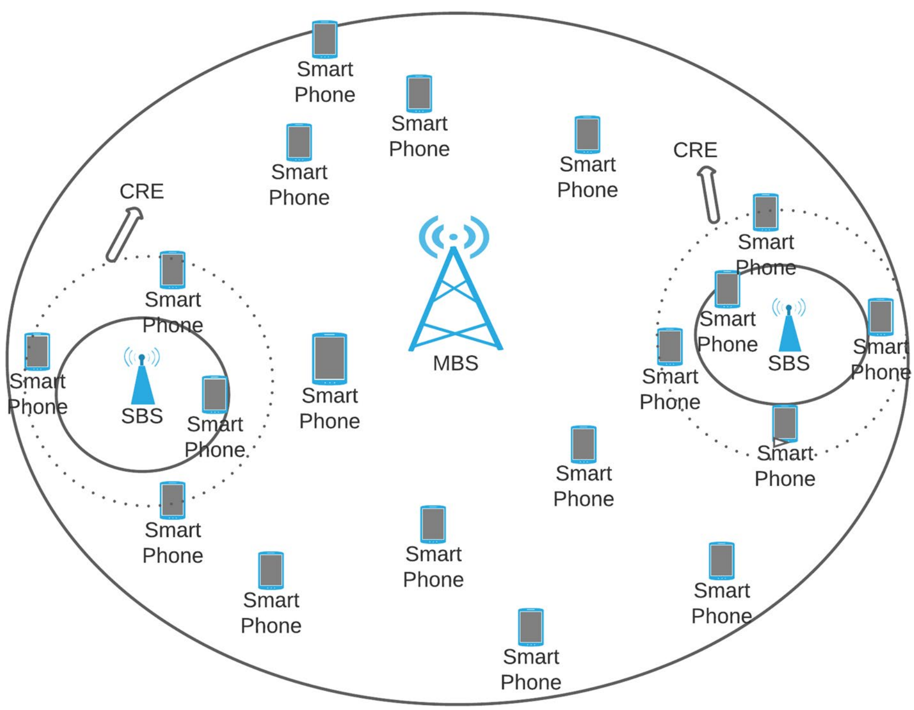


**Fig. 1** Heterogeneous network indicating cell range extension

reference signal received power (RSRP), user equipments (UEs) typically connect with the cell possessing the highest received downlink power. Downlink power refers to the power level of the signals transmitted from the MBSs or SBSs to the UEs, i.e., the mobile devices [9]. This distribution of UEs throughout the network becomes uneven, with UEs predominantly concentrated on the MBSs [10].

Load balancing in wireless networks refers to the even distribution of traffic, data, and computational tasks across network components, such as base stations or access points, to prevent congestion and ensure optimal resource utilization. Load imbalance can lead to several issues, including reduced network capacity, uneven resource utilization, decreased user experience, and increased energy consumption. In the realm of mobile cellular networks, the term 'load' encompasses various interpretations, often linked to the count of UEs connected to a base station (BS). When the aggregate number of resource blocks (RBs) allocated to users remains constant, the traffic load encountered by the base station corresponds directly to the number of attached UEs [11, 12]. In 5G and beyond HetNets, load balancing becomes more critical due to the diverse range of devices, services, and applications these networks support. Hence, load balancing assumes a pivotal role in the context of HetNets. The method governing the association of BSs and users must be optimized to maximize the utilization of system resources to deliver an enhanced QoS to end users. To address these challenges, load balancing mechanisms aim to dynamically allocate users and traffic to different network components based on factors such as signal strength, channel conditions, and network capacity so that traffic is evenly distributed across various network components. Simultaneously, interference mitigation is essential to ensure uninterrupted and reliable communication, particularly in scenarios where multiple devices and users coexist in the same environment [13]. Interference in wireless networks occurs when signals from different devices or base stations overlap, leading to performance degradation and reduced data rates. Interference is particularly more problematic in densely populated areas and scenarios where multiple devices transmit simultaneously. Traditional solutions to these challenges often involve intricate trade-offs and compromises to network performance [14].

The standardized technique established by the 3GPP, known as cell range extension (CRE), emerges as a promising avenue to attain improved equilibrium among users within a mobile network. CRE involves assigning a bias factor to each SBS, effectively modifying the coverage area's extent (as depicted in Fig. 1), thereby rendering users more inclined to connect with a given SBS. This strategy enables optimizing radio resource utilization through a more widespread distribution of users across the mobile network. Addressing the challenges inherent to CRE implementation involves precise bias value computation for each layer within the HetNet [15] or even determining a tailored bias value for individual SBSs. These methodologies predominantly rely on resolving combinatorial optimization problems [16] by maximizing the cumulative data link rates for users. However, these approaches often overlook the individual needs of users when considering CRE integration. While the pursuit of better system serviceability stands as a valid strategy, a more effective solution could be developed through the adoption of application-aware strategies to meet users' specific traffic requirements.

Furthermore, the increasing demand for mobile data is making it harder to manage mobile networks effectively. While many methods have been proposed to improve network performance, the real challenge is dealing with the complexity of mobile networks using methods inspired by nature and artificial intelligence. This is a growing field of research that combines ideas from machine learning, bio-inspired algorithms, and fuzzy neural networks [17]. These methods are used to optimize computer systems in various situations. Bio-inspired algorithms, which learn from the behavior of species in nature, offer a way to handle complex systems efficiently. These methods can potentially improve the design, maintenance, and optimization of self-organized networks (SONs) [18]. Notably, these techniques maintain a relatively modest level of complexity, facilitated by local interactions and iterative feedback-based learning. Hence, combining bio-inspired mechanisms with network operation strategies presents a promising avenue for enhancing user association and load balancing processes within a heterogeneous cellular network.

Moreover, significant emphasis has been placed on leveraging UE data traffic demands to establish customized bias values for individual BSs. While striving to distribute the load uniformly across network layers, there's a risk of certain BSs becoming either excessively overloaded or underutilized. Consequently, the imperative lies in achieving load equilibrium for each BS, along with synchronized real-time network resource optimization. This approach is aligned with the seamless operation of mobile networks, without necessitating supplementary signaling mechanisms. The ultimate goal is to enhance the network's capacity to cater to UE devices' QoS and traffic prerequisites.

In this context, the application of multi-objective optimization techniques offers a promising avenue to overcome the complexities of load balancing and user association simultaneously. Ant colony optimization (ACO) [19, 20], inspired by the behavior of ants in nature, has proven its efficacy in solving intricate combinatorial problems. By leveraging ACO, we aim to develop a comprehensive approach that optimizes both load distribution and cell association in 5G and beyond HetNets.

This paper delves into the novel fusion of multi-objective optimization and ACO for tackling the intricate challenges of load balancing and user association in 5G and beyond HetNets. The proposed approach strives to strike a delicate balance between efficient resource utilization, seamless user experience, and the evolving demands of modern communication scenarios. By exploring this innovative approach, we endeavor to contribute to the advancement of network optimization strategies that pave the way for enhanced connectivity in the era of 5G and beyond. Hence, the significant contributions of this paper are briefed below:

1. We formulate a strategy involving CRE, integrating it with the ACO algorithm to account for the distinct traffic demands of mobile users and achieve enhanced control over network load distribution. The utilization of the ACO algorithm as an optimization mechanism aims to calculate the bias values for all SBSs collectively, striving for a unified approach.
2. We construct a user association problem to maximize the count of UEs that satisfy their downlink prerequisites. This approach differs from existing studies by primarily focusing on maximizing the throughput rate as well as minimizing the call drop rate for mobile users.

3. We extensively examine the performance of the objective function across various bias values and explore the impact of parameter combinations on both the user association problem and the equilibrium of network load distribution.
4. By conducting thorough simulation experiments within a two-tier HetNet using MATLAB simulator, we measure the effectiveness of our approach and conduct a comparative analysis against recent relevant research. The outcomes reveal that our proposed method holds significant potential for 5G and beyond ultra-dense HetNets, as it demonstrates superior performance in terms of throughput, load balancing, and call drop rate within HetNets.

The primary contribution of this paper lies in its significance for the telecommunications sector and the end users of smart devices, facilitating the utilization of advanced networks like 5G and beyond. The telecommunications industry is currently advancing toward the development of 6G networks, expected to be launched by 2030. However, several critical challenges must be addressed to achieve this ambitious goal, including issues such as user association, load balancing, interference reduction, and resource management. Consequently, the principal objective of this research paper is to address and resolve the challenges associated with user association and load balancing.

The subsequent sections of this paper are structured as follows. Section 2 delves into related research on user association and load balancing strategies. Section 3 discusses the system model, scenario implementation, and incorporation of decision variables. Section 4 outlines the problem formulation. In Sect. 5, the proposed methodology is explained. In Sect. 6, we provide details of the experiments conducted and present the simulation results. Finally, Sect. 7 ends with concluding remarks and a discussion on potential future directions.

## 2 Related works

This section delivers a comprehensive survey of literature pertaining to user association along with load balancing strategies and the integration of bio-inspired strategies within heterogeneous mobile networks.

Given the evolving landscape of large-scale, heterogeneous, dynamic, and intricate scenarios in HetNets, numerous studies have been undertaken to assess the efficacy of computational intelligence techniques in enhancing the operational performance of HetNets [21]. References [21, 22] reviewed various AI-driven methodologies to foster the evolution of more intelligent HetNet infrastructure and systems. This review particularly addressed the research related to self-configuration and self-optimization, while providing a balanced analysis of the advantages and drawbacks inherent in each of these associated schemes. Furthermore, the study referenced specific instances of particle swarm optimization (PSO) strategies for enhanced user association within cellular networks. The authors in [21] suggest that as technology in HetNet advances, we need new ways to meet the growing needs of the Internet of Things (IoT) and machine-to-machine (M2M) services. They propose using smart optimization and management methods inspired by how nature's ecosystems work. Specifically, they recommend combining techniques like genetic algorithms (GAs) and ACO from swarm intelligence to handle these needs effectively.

The comprehensive examination of the application of swarm intelligence in communication networks is thoroughly elaborated in [23]. The paper looks at different bio-inspired techniques and algorithms to make artificial systems work better. It emphasizes the critical attributes of swarm intelligence observable in social species, thus delving into intelligent traits like adaptability, resilience, decentralized control, and self-evolution. These traits are subsequently explored for their potential applicability within communication networks. Furthermore, the paper also outlines how these techniques can help with designing wireless communication systems, making them work better, and optimizing how they function.

Reference [24] focused on the application of bio-inspired techniques, wherein several algorithms relevant to artificial SONs were explored. The authors highlighted several unresolved research questions, encompassing issues like SON design trade-offs, the incorporation of optimization capabilities within LTE-A systems, cross-layer design, cognitive self-optimization for M2M scenarios, resource allocation, and power regulation. Moreover, the paper provided insights into several critical issues of SONs involving media access control (MAC)-layer and physical-layer operations. The paper asserted that swarm intelligence holds promise as an effective technique for SON management. However, apart from SONs, the bio-inspired techniques that can be universally adapted across diverse network environments remain a formidable challenge.

The authors in [25] proposed an algorithm named P5G, based on particle swarm optimization (PSO), aimed at optimizing various key performance indicators (KPIs) through the allocation of virtual components termed as reusable functional blocks (RFBs) within the domain of software-defined networks (SDNs). P5G seeks to efficiently manage RFBs for seamless delivery of HD videos to users, within a standard scenario involving SBSs, MBSs, and evolved packet core (EPC) nodes. Despite omitting user-specific traffic requirements, the outcomes demonstrate that P5G closely approximates the optimal solution, with consistently low computation times.

In [26], PSO is harnessed to dynamically determine biasing values for BSs with the objective of optimizing the attainable throughput. The system framework revolves around a three-tier downlink HetNet configuration, wherein distinct path loss models are applied for each category of base station. The proposed strategy demonstrated the capability to ensure equitable treatment of users by regulating the load distribution among BSs. However, it is important to note that the movement of particles within the search space is unaffected by user-specific traffic requisites.

The study undertaken in [27] examines the application of an algorithm embodying a Bayesian base station selection game to address the user association problem. This methodology considers the distinctive attributes of SBSs and the nature of user traffic requirements, aiming to improve the likelihood of suitable associations that minimizes the end-to-end latency experienced by UEs. The proposed strategy is validated through metrics such as the probability of accurate association, latency improvement compared to conventional CRE, and the effectiveness of user association algorithms in terms of SINR inherent in LTE-Advanced. Notably, as this investigation focused on a scenario involving a sparse distribution of base stations, the technique's feasibility could be inconclusive due to the utilization of combinatorial optimization schemes. The investigation outlined in [28] employs stochastic geometry methodologies to dissect the user association challenge, with the objective of identifying optimal bias values that drive the augmentation of throughput achieved by UEs. Furthermore, this paper introduces bias factors in a mathematically determined expression to comprehensively assess distinct network configurations on the virtual coverage regions. Nevertheless, it is important to note that the analysis within this research does not account for the precise attributes of user traffic, which could potentially enhance the precision of the user association mechanism.

Reference [29] proposed a method to concurrently address cell selection and resource allocation with the objective to alleviate the burden on MBSs by freeing up a substantial number of resource blocks (RBs) and directing users toward SBSs. Simulation outcomes indicate that the proposed framework potentially outperforms the state-of-the-art CRE schemes. However, the method has a significant drawback in that it relies on a collection of non-standardized signals, potentially leading to an increased overhead level. This particular characteristic renders the proposal less suitable for scenarios involving UDNs. The study in [30] introduces a method for load balancing and user association, underpinned by the knapsack optimization (KO) algorithm. The primary objective is to achieve a balanced distribution of users across various layers of SBSs. The approach is subject to constraints linked to the service capacity of BSs and the volume of resource blocks required by users to fulfill their data rate stipulations. It is noteworthy that as the count of BSs and UEs escalates, such as in UDN, the solution might encounter limitations in terms of scalability and the time taken to converge. This is due to the fact that KO problems are NP-hard, thereby affecting the feasibility within multi-tier HetNets.

Velmurugan et al. [31] aimed at reducing the frequency of hand-off calls and power consumption of mobile nodes by using the multi-objective whale optimization (M-WO) algorithm for vertical hand-off (VHO) decision making. Simulation results showed that under certain conditions, the multi-objective whale optimization (M-WO) algorithm outperformed some existing algorithms such as optimization based on SSF (strongest signal first) and OPTG, in terms of call drop rate, load, and battery lifetime of the mobile terminal. In [32], the authors formulated a joint optimization problem of user association, channel allocation, and reuse pattern selection. The primary strategy is to ensure that UEs align with the closest SBSs. This is based on the condition that the SINR received surpasses a pre-established threshold. If this criterion is not met, UEs are inclined to associate with MBSs. By adjusting the threshold, the paper demonstrates that this rule yields notably enhanced performance when compared to the SINR association technique (in the absence of cell range expansion).

Umar et al. [33] presented innovative hybrid multiple access methods like phone user (PU) clustering-based downlink in HetNets to optimize throughput and QoS by enhancing the cell association. Using an outer approximation algorithm, it addresses the mixed integer nonlinear programming optimization problem derived from the concave fractional programming problem. Results indicate its superiority over macrocell-only networks and HetNets in terms of cell association, throughput, and QoS. Maryam et al. [34] addressed user association challenges in backhaul-constrained HetNets, emphasizing the importance of load balancing for optimal small cell

**Table 1** Parameters used in the simulation

| Simulation trial number | 1000 |
|---|---|
| X | Refers to user positions |
| O | Refers to BS positions |
| MBS number | 1 |
| Area covered by MBS | 1 $km^2$ |
| MBS radius | 500 m |
| PBS number | 3 |
| PBS radius | 50 m |
| Total No. of sub-channels, K | 100 |
| Each sub-channel bandwidth | 180 KHz |
| MBS power transmission | 43 dBm |
| SBS power transmission | 23 dBm |
| Carrier frequency | 2 GHZ |
| Channel | Rayleigh fading |
| Environment | Urban |
| Noise power spectral density | − 174 dBm/Hz |
| User noise figure | 9 dB |
| Antenna gain | 5 dB |
| Pico cell path loss model | 140.7 + 36.7 * log10(*d*) dB (*d* [Km]) |
| Macro cell path loss model | 128.1 + 36.7 * log10(*d*) dB (*d* [Km]) |
| Shadowing standard deviation | 10 dB |

utilization. It underscores the need for a combined approach of cell association and interference management to safeguard offloaded users from detrimental interference.

Moreover, a majority of the previously discussed studies are confined to employing Shannon's theorem, in contrast to the conventional 3GPP-LTE discrete modulation-and-coding scheme (MCS) function. This distinction can potentially lead to an overly optimistic estimation of network capacity. Utilizing a discrete MCS function, however, introduces notable complexity when determining the optimal association for cell-edge users. This complexity arises due to the absence of the convexity and strictly ascending attributes inherent in Shannon's theorem [35].

## 3 System model

In this section, we illustrate the system model and develop a formulation for an optimization problem tailored to compute distinct CRE bias values to be allocated to each SBS, considering the specific requirements of users' traffic patterns. The parameters used for simulation are summarized in Table 1. We assume a downlink HetNet configuration, encompassing $K$ autonomous network tiers of base stations where $K$ ranges from 1 to k. Within this setup, a typical UE is positioned at location $\mathbb{R}^3$. The arrangement involves the sampling of user and BS positions through separate distributions derived from homogeneous Poisson point processes (HPPP). This means, the positions of the BSs and users are chosen randomly according to HPPP but one fixed case of that scenario has been used as illustrated in the simulated system model in Fig. 2. In this context, each $k$th layer exhibits a density of $\lambda_k$, with its BSs being randomly generated via HPPP $\phi(\lambda_k)$, while user positioning is accomplished through HPPP $\phi(\lambda_u)$.

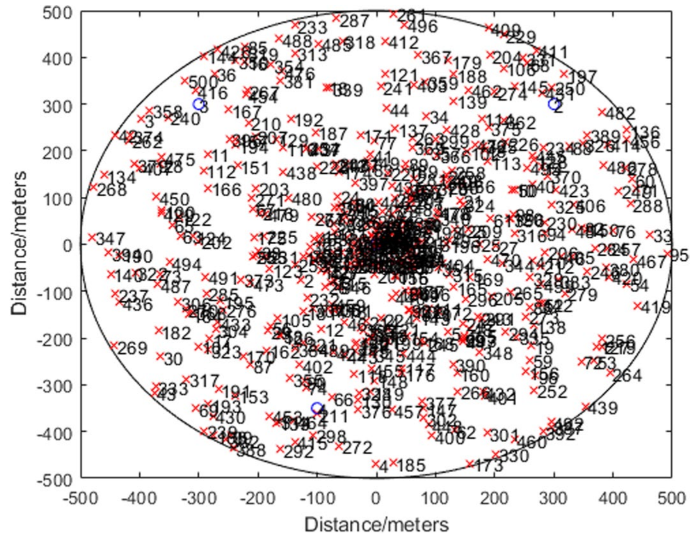


**Fig. 2** Simulated system model

The collection of all base stations is stated as $\phi$, where $\phi = (\delta \cup \gamma)$. The group of MBSs is denoted by $\delta = M1, M2, M3, \ldots, Mm$, while the SBSs are represented by $\gamma = S1, S2, S3, \ldots, Ss$. Here, $\phi$ is indexed with $1 \leq j \leq b$ (where $b = m + s$). The set of UEs is symbolized as $\pi = U1, U2, U3, \ldots, Uu$, where $1 \leq i \leq u$. Additionally, $\psi = \theta_1, \theta_2, \theta_3, \ldots, \theta_s$ represents the set of bias values associated with SBSs. Furthermore, the $i$th user requests a specific service class represented by the tuple $\rho_i = (\eta_i, \tau_i)$, wherein $\eta_i$ and $\tau_i$, respectively, indicate the average throughput and the compression factor. Consequently, the required data rate for the $i$th UE can be computed as the product of $(\eta_i \cdot \tau_i)$. CRE for Max-SINR is formulated in the following paragraphs.

User association techniques rely on various metrics, including reference signal received quality (RSRQ), RSRP, and SINR. Among these, RSRP and RSRQ stand out due to their minimal extra communication complexity, as these parameters are already defined in LTE standards. A study conducted by [36] assesses the user association mechanism using these metrics and demonstrates that selecting based on SINR can result in improved downlink rates. Therefore, in this model, we posit that the received SINR serves as a pivotal indicator of the UE rate and outage performance, given its direct correlation with Shannon's theorem. Adhering to the Max-SINR association criteria, the $i$th UE inclines toward associating with the $j$th BS, where $j = \arg\max(\text{SINR}_{ij})$ for all $j \in \phi$. In the context of a two-tier HetNet ($K = 2$) and the prevailing situation that MBSs possess considerably greater transmission power ($P1 \gg P2$), UEs exhibit a propensity to primarily associate with MBSs. By introducing a CRE bias to the SINR of each SBS, UEs experience enhanced distribution among BSs, potentially leading to an improved long-term rate for each UE. For instances where the $i$th UE gravitates toward the MBS tier (tier-1) and selects an MBS $k \in \delta$, the received SINR ($\zeta$) adheres to conditions (1) and (2):

$$\zeta_{ik} > \zeta_{ij}, \quad \forall j \in \delta, \quad k \neq j. \tag{1}$$

$$\zeta_{ik} > \zeta_{ij} + \theta_j, \quad \forall j \in \gamma. \tag{2}$$

In this context, $\theta_j$ denotes the CRE bias assigned to the SBS indexed as $j$. Furthermore, with the implementation of CRE, the $i$th UE opts for the $\ell$th SBS when the received SINR fulfills the conditions specified by Eqs. (3) and (4):

$$(\zeta_{i\ell} + \theta_\ell) > \zeta_{ij}, \forall j \in \delta. \tag{3}$$

$$(\zeta_{i\ell} + \theta_\ell) > (\zeta_{ij} + \theta_j), \forall j \in \gamma, j \neq \ell. \tag{4}$$

By adjusting the appropriate bias or cell specific offset (CSO) values for the small BSs as explained earlier, these base stations modify their coverage area for downlink, leading to an increase or decrease in the number of users associated with them. When the $i$th user is linked to the $j$th base station, the downlink SINR ($\zeta_{ij}$) can be formulated as:

$$\zeta_{ij} = P_i G_{ij} / \left( \sum_{k \epsilon \phi} P_k G_{ik} + P_{AWGN} \right) \tag{5}$$

In Eq. (5), $P_i$ is stated as the power transmitted by BS $s$, $G_{ij}$ is the gain of the downlink channel for the link of BS $s$ and UE $u$, and $P_{AWGN}$ is the additive white Gaussian noise power. The power transmitted by the interfering cells is denoted as $P_i$ and $G_{ij}$ is their gain. Since cell selection is considered to be run on a larger time scale, $G_{ij}$ is deemed to have been averaged within the period of association and overall physical resource blocks in the whole channel spectrum, which means, frequency-selective fading and fast fading are averaged out. Hence, $G_{ij}$ remains constant despite the dynamic channel variations within the cell selection period, and the SINR between base station $j$ and user $i$ for each sub-channel is similar. Therefore, the attainable per-channel downlink rate for the $i$th user connected to the $j$th base station can be formulated as:

$$R_i = e_\ell \cdot \frac{n_{\text{sc}} \cdot n_{\text{sym}}}{T_{\text{subframe}}} \tag{6}$$

Here, $e_\ell$ signifies the efficiency per sub-carrier with respect to bits for each orthogonal frequency-division multiplexing (OFDM) symbol at a designated threshold SINR. This threshold SINR is typically chosen based on various factors such as the desired quality of service (QoS), system requirements, and environmental conditions. One limitation of our simulation study is that it assumes a constant SINR threshold, whereas, in practical scenarios, this threshold varies. However, the SINR threshold has been the same for all the compared algorithms. This efficiency, denoted as $e_\ell$, is derived through the utilization of an MCS function, represented as $\mu(\zeta_{ij})$. The MCS function maps the SINR values to specific modulation and coding schemes, which determine the achievable data rate and reliability of communication over the sub-channel. The variables $n_{\text{sc}}$, $n_{\text{sym}}$, and $T_{\text{subframe}}$ correspond to the count of subcarriers per channel, the quantity of OFDM symbols within each subframe, and the time length of each subframe, respectively. Through the adoption of an equitable resource allocation strategy, wherein the total resource blocks (RBs) are evenly distributed among the linked users; the complete number of resource blocks attained by the $i$th user from the $j$th BS can be mathematically formulated as:

$$n_{\mathrm{RB}_i} = \frac{n_{\mathrm{RB}_j}}{L_j} \quad (7)$$

where $n_{\mathrm{RB}_j}$ refers to the total number of available RBs in base station j; meanwhile, $L_j$ denotes the total load of users associated with base station j. A similar type of model has been implemented in [37], which has been adopted in this paper as the equations are common to best suit our proposed ACO algorithm.

## 4 Problem formulation

Let us assume that the matrix denoted as $X$ is a binary matrix that indicates the connection between the $i$th user and the $j$th base station. Consequently, the value of the factor $x_{ij}$ is determined by:

$$x_{ij} = \begin{cases} 1 & \text{if the } i\text{th user is linked up to the } j\text{th BS;} \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Moreover, assume that $Y$ signifies an array composed of binary values that symbolize the satisfaction of the user's demand regarding the downlink rate. Specifically, the element $y_i = 1$ if the $i$th user's download needs are fulfilled, as depicted by:

$$y_i = \begin{cases} 1 & \text{if the } i\text{th user's downlink demand is satisfied;} \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

Furthermore, assume that $Z$ denotes an array that contains binary values. The factor $z_j$ represents whether the $j$th base station is associated with at least one user, as demonstrated below:

$$z_j = \begin{cases} 1 & \text{if the BS j is associated with at least one user;} \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

In this section, taking into account the performance criteria and decision variables employed in this study, we establish an optimization problem characterized by the objective function presented below:

$$\text{Maximize } \alpha \cdot \sum_{i \in \pi} y_i + \beta \cdot \sum_{j \in \phi} z_j \quad (11)$$

The aim of Eq. (11) is to enhance the satisfaction of users by maximizing the count of UEs whose downlink needs are fulfilled and the count of base stations that are linked to UEs. When the element $z_j$ is maximized, it is anticipated that both the number of base stations connected to UEs and the number of resource blocks utilized by those users will increase. The elements $\alpha$ and $\beta$ harmonize the involvements of the different elements in the objective function. Moreover, the maximization goal is tied to the selection of predefined values $\psi = \{\theta_1, \theta_2, \theta_3, \ldots, \theta_s\}$, which directly impact the values derived from $y_i$ and $z_j$. Furthermore, the objective function is bound by the following constraints:

$$\sum_{j \in \phi} x_{ij} = 1, \quad \forall i \in \pi, \quad (12)$$

$$\sum_{i\in\pi} n_{\mathrm{RB}_i} \leq n_{\mathrm{RB}_j}, \quad \forall j \in \phi, \tag{13}$$

$$n_{\mathrm{RB}_i} \geq T_{\mathrm{B}}, \quad \forall i \in \pi, \tag{14}$$

Constraint (12) ensures that each user is linked with just one base station, implying that a Coordinated Multipoint Transmission (CoMP) is not taken into account. The following constraint (13) safeguards that the RBs utilized by the $i$th UE do not exceed the total RBs accessible at the $j$th base station. Lastly, constraint (14) assures solution feasibility by requiring that the resource blocks received by a user must exceed a specified minimum threshold $T_{\mathrm{B}}$.

## 5 Proposed methodology

This section describes the simulation parameters for the experiment. The ACO algorithm used for optimizing fairness and throughput is also discussed here.

### 5.1 Mobile network configuration and parameters

The proposed methodology uses the simulation parameters as provided in Table 1, referenced from 3GPP, and the path loss model followed is distance-based. It is ensured that similar efforts are given to all the parameters of each optimization methods. The simulation involves the evaluation of a two-tier HetNet, with $K = 2$, where 1 MBS and 3 SBSs are used to represent a UDN scenario.

Furthermore, the transmission power levels are set at 46 dBm and 23 dBm for MBS and SBSs, respectively. We conduct a total of 1000 trials for each scenario. The locations of users and base stations are generated based on their respective HPPP densities, covering an area of 1 km$^2$. Subsequently, the SINR for potential links is determined and connections between UEs and their corresponding base stations are established via the Max-SINR association scheme. The SINR computations facilitate the determination of the downlink rates for UEs. The downlink SINR is derived using Eq. (8), with Max-SINR representing the maximum value of this SINR. Under the Max-SINR association scheme, a user is associated with a specific BS if its downlink SINR surpasses that of the second strongest signal; otherwise, it is associated with the next best Max-SINR BS. Additionally, the downlink rate is computed using Eq. (6).

### 5.2 The ACO algorithm

In this experiment, we introduce the ACO algorithm for the computation of individualized CRE bias or also known as CSO for each SBS. The objective is to determine unique CSO values, which correspond to the elements within the set $\psi = \{\theta_1, \theta_2, \theta_3, \ldots, \theta_s\}$. In a natural setting, ants use information about pheromone concentration to find the shortest path between their nest and food source [38]. This ant behavior serves as the inspiration for the bio-inspired ACO algorithm. The fundamental concept behind employing this CRE algorithm based on ACO is to parallelize the search process across multiple concurrent computational threads. This approach utilizes a dynamic memory structure that incorporates information regarding the previous actions' success by computational agents. This information provides guidance for the construction process of ants in subsequent cycles or iterations. The central component of this algorithm is the artificial ant, which serves as a relatively straightforward computational agent responsible for probabilistically and iteratively constructing the solution to the current problem. The intermediary partial solutions created during this process are referred to as states. Each ant transitions from one state to another, gradually constructing a complete solution. At each step, every ant calculates a set of feasible extensions to the current state and probabilistically selects one of them to create a new state. The probability of an ant moving from state $s$ to state $s+1$ is influenced by two key factors:

1. The initial preference for transitioning from state $s$ to state $s+1$, which is determined by a problem-specific heuristic.
2. A measure of the effectiveness of a specific move in the past for constructing an optimal solution.

These move-related indicators, referred to as trails, undergo updates at the conclusion of each iteration. This involves increasing the importance of moves that contribute to optimal or near-optimal solutions while decreasing the significance of other moves. As discussed in the introduction section, the primary goal of this research is to achieve efficient load balancing by effectively redistributing UEs between macro and Pico e node base stations (eNBS). To accomplish this, we have followed the below methodology:

Let $N$ represent the number of UEs randomly distributed in the network, which includes 1 Macro eNBS and $p$ Pico eNBs, where in this case $p = 3$. We introduce $n$ as the attribute based on which UEs need to be grouped into clusters, in our case which is SINR. Consider $K$ as the number of clusters. We use $R$ to denote the agent responsible for constructing the solution, and $S$ represents the solution string to be generated by the algorithm. The pheromone matrix has a size of $N \times K$, where it denotes the concentration of pheromone at the location of $UE_{i=1:N}$ associated with cluster $j = 1, \ldots, K$. To produce a solution $S$, the agent determines the cluster number for each element of the string $S$ via one of the following methods:

- *Exploitation process* The cluster with the highest pheromone concentration is chosen based on a probability $0 < \alpha < 1$.
- *Biased exploration process* One of the $K$ clusters is selected by utilizing the stochastic distribution with a probability $1 - \alpha$, denoted as $P_{ij}$, which represents the normalized pheromone probability for parameter $i$ which is inside the cluster $j$ can be expressed as below,

$$P_{ij} = \frac{[\mathcal{T}_{ij}]^{\alpha}[n_{ij}]^{\beta}}{\sum [\mathcal{T}]^{\alpha}[n_{ij}]^{\beta}} \tag{15}$$

Here, $\mathcal{T}_{ij}$ is the pheromone matrix that determines the amount of pheromone present in between the food and nest. The parameters $\alpha$ and $\beta$ represent the relative importance of the pheromone, and $n_{ij}$ is the inverse of the distance between $i$ and $j$, i.e., food and nest.

The objective function evaluates the solution's quality by summing the squared Euclidean distances between the center of the cluster and each object inside the cluster. When dealing with $N$ UEs $\{x_1, x_2, \ldots, x_n\}$ that need to be grouped into $K$ clusters, the mathematical representation of this problem regarding the clustering of data can be defined as the minimization of the objective function $F(w, m)$:

$$F(w,m) = \sum_{j=1}^{K} \sum_{i=1}^{N} \sum_{v=1}^{n} w_{ij} \left\| x_{iv} - m_{jv} \right\|^2 \tag{16}$$

Here, $w_{ij}$ represents the weight associated with UE $i$ inside the cluster $j$. $w_{ij}$ is a binary weight variable that equals 1 if object $i$ is inside the cluster $j$, and it equals 0 otherwise. $x_{iv}$ denotes the $v$-th attribute of UE $i$, and $m_{jv}$ signifies the $v$-th attribute of the center of cluster $j$. The aim of this is to minimize this function.

In every loop, individual ants generate solutions. To enhance these solutions, a straightforward local search method is applied to the best solutions with the lowest values of the objective function (referring to the highest fitness values). The cluster assignments of every UE sample in the solution string are modified with a definitive threshold probability. These newly generated solutions are accepted only if they result in improved fitness. Subsequently, the pheromone matrix is updated using the following rule:

$$\tau_{ij}(t+1) = (1-\rho) \cdot \tau_{ij}(t) + \Delta\tau_{ij}, \quad \text{for } i = 1, \ldots, N, \text{ and } j = 1, \ldots, K \tag{17}$$

Here, $\rho$ represents the trail persistence, with a range between 0 and 1, and $(1-\rho)$ denotes the evaporation rate. An optimal solution is the one that minimizes the objective function value. The algorithm stores the solution associated with the minimum objective function value in memory and continues iterating. Upon reaching the stopping criterion, the best solution, characterized by the lowest objective function value, becomes the final solution. Thus, the best possible bias for CRE is sorted out via the ACO technique. A flowchart is presented in Fig. 3 to show how the ACO algorithm works.

The steps of sorting out the best bias value can be understood from the flowchart in Fig. 3. Initially, the ant colony optimization algorithm is initialized, setting parameters like the number of ants, iterations, pheromone evaporation rate, and initial bias values. Subsequently, ants are dispatched to explore the solution space, representing potential solutions to the problem of bias value selection. Their movement encompasses the space of potential bias values for extending cell range. During exploration, the encountered bias values are assessed for quality based on predefined objectives, such as minimizing interference and maximizing load balance. Pheromones are deposited by ants on their paths, facilitating communication and information sharing among them. The amount of pheromone deposited corresponds to the quality of the encountered bias value, which is continually updated according to the fitness function's findings. Following the ants' exploration, a global update is performed to adjust pheromone levels based on the solutions found collectively by the ant colony. This process strengthens paths leading to favorable bias values while weakening those leading to sub-optimal solutions. Iterations continue through ant movement, bias evaluation, and pheromone update, gradually converging the ant colony toward the optimal bias values for cell range extension. The process concludes upon finding the optimal value that satisfies the desired objectives, signaling the cessation of iterations.

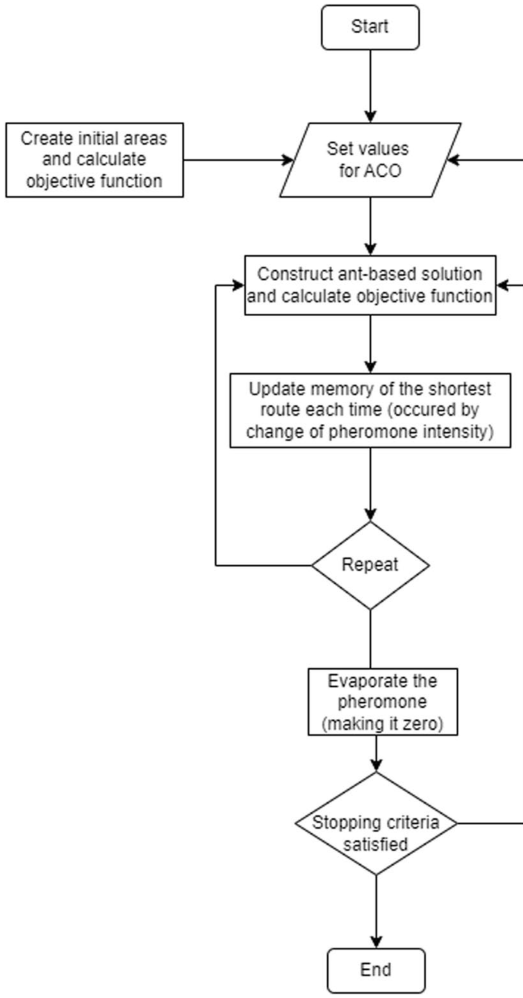


**Fig. 3** Flowchart of how proper bias is sorted out by ACO

**Table 2** ACO parameter values

| Description of parameters | Value |
|---|---|
| Number of total UEs | 500 |
| Number of cycles | 100 |
| Number of ants | 100 |
| Initial pheromone level | 0.05 |
| Pheromone persistence rate | 0.5 |
| Probability of choosing next node | 0.75 |
| Number of parallel execution | 4 |
| Feedback window size | 10 |

Table 2 represents the parameters selected for the ACO algorithm. ACO incorporates specific parameters like pheromone persistence rate and its initial value, as well as the number of parallel executions, which enhance its effectiveness in determining optimal bias values in scenarios akin to this study. On the other hand, PSO involves unique parameters such as inertia weight, cognitive weight, and inertia weight decay. Similarly, the remaining compared algorithms, M-OPTG and M-WO, possess their distinct parameters. In this investigation, ACO has been implemented and compared with the results obtained from similar studies utilizing these alternative techniques (PSO, M-OPTG, and M-WO). While it is plausible that another algorithm might outperform if optimized with superior parameters, the parameters employed are assumed to remain consistent throughout this study.

## 6 Results and discussion

In the following section, the simulation results are thoroughly discussed and analyzed to understand how the proposed ACO-based CRE methodology outperforms the state-of-the-art techniques such as CRE based on M-OPTG, M-WO, and PSO. In the simulations, the parameters under comparison include Jain's fairness index (JFI), which assesses load balance, as well as throughput and average throughput, determined using the conventional formula derived from Eq. (6), and call drop rate. Throughout the simulation process, all parameters are kept consistent across the algorithms being evaluated to ensure uniform calculation standards.

Figure 4a, b illustrates the proposed ACO-based CRE method's performance in load balancing and throughput for varying numbers of users in two different sub-plots

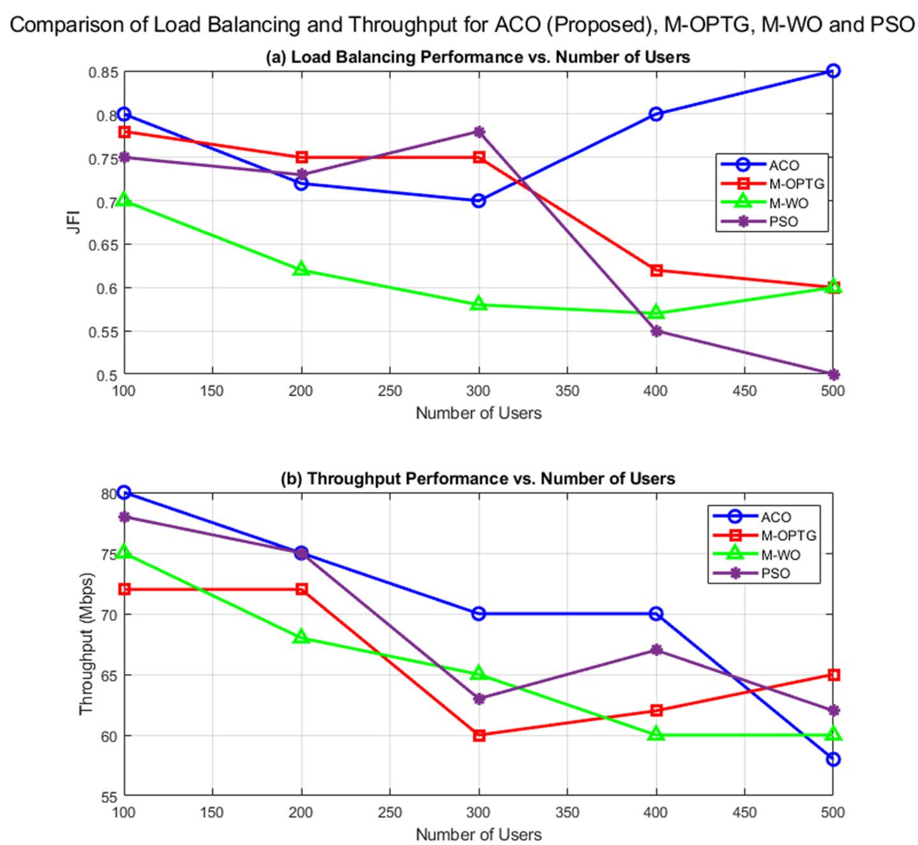


**Fig. 4** **a** Comparison of load balancing and **b** throughput for ACO (proposed) and other state-of-the-art techniques over variable users

simultaneously. Load balancing is assessed using Jain's JFI, a parameter commonly used by researchers in the same domain. As illustrated in Fig. 4a, b, even with the increase in user numbers, the proposed ACO-based CRE method consistently exhibits superior performance in terms of both JFI and throughput compared to M-OPTG, M-WO, and PSO. It has been observed that there has been a downward trend for throughput while it is an upward trend for JFI, specifically while the number of users is 300–500. This suggests that improving the load balancing of the network appears to occur at the expense of the network's throughput, particularly when there is a higher density of users.

Figure 5 depicts the network performance of the proposed ACO-based CRE method in terms of average throughput for varying numbers of hand-offs in comparison to M-OPTG, M-WO, and PSO. In general, all the techniques exhibit an upward trend as the number of hand-offs rises. However, the proposed ACO-based CRE method demonstrates superior performance. This suggests that the proposed ACO-based CRE method would demonstrate similar superior performance with an even greater number of small cells, as hand-offs typically occur when users transition from one cell area to another. During a hand-off, a user's device transitions from one network access point to another. This process involves briefly disconnecting from the current network and reconnecting to the new one. While the disconnection is usually short, it can temporary interrupt data transmission, impacting the user's experience. The extent to which throughput decreases during a hand-off depends on various factors, including the efficiency of the hand-off mechanism, the quality of the new network connection, and the speed at which the user's device can establish the new link. In the simulations, we have assumed ideal conditions where the hand-off process is seamless and the new network connection is of higher quality, so the impact on throughput may be minimal. However, in real-world scenarios, factors such as network congestion, signal interference, and device capabilities can contribute to a noticeable decrease in throughput during hand-offs. However, an

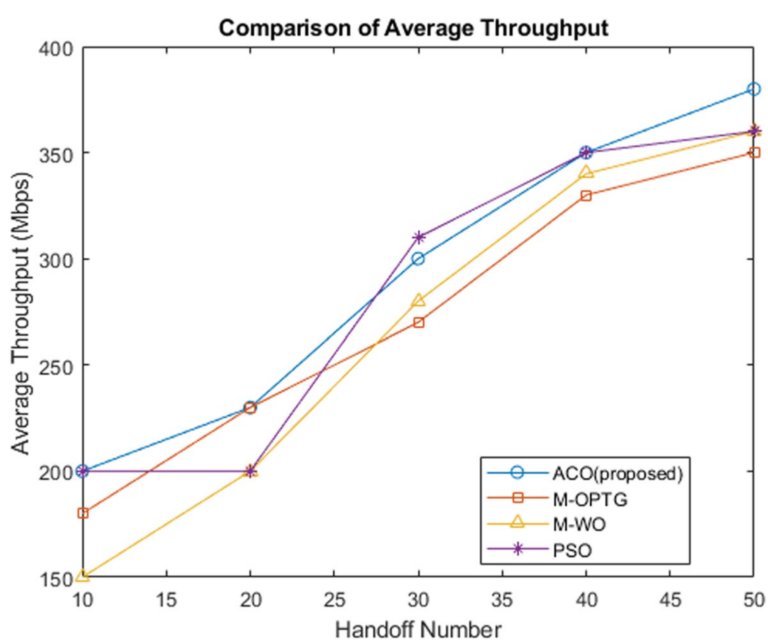


**Fig. 5** Comparison of average throughput of all BSs over variable hand-off number

efficient hand-off mechanism allows for an increase in the average throughput. Likewise, the load balancing performance is notably better for the proposed ACO-based CRE method in comparison to its closest competitor, PSO, as depicted in Fig. 6. Conversely, the M-OPTG and M-WO based CRE methods consistently under-perform, as shown in Fig. 6. This proves that ACO is inherently better at exploring the solution space globally compared to PSO, M-OPTG, and M-WO. This is because ACO employs multiple ants exploring different regions simultaneously, leading to a more comprehensive search for optimal bias values. This global search capability helps ensure that the solution found is not trapped in local optima, enhancing the chances of discovering superior solutions for load balancing.

Figure 7a presents an analysis of the call drop rate with varying numbers of hand-offs as well as Fig. 7b presents the comparison against varying numbers of users. The analysis of call drop rate aims to ensure that nearly uninterrupted service is provided to the users while simultaneously improving user throughput and load balance of the BSs. It is observed from Fig. 7a, b that the proposed ACO-based CRE method demonstrates better overall performance with the increase in the number of hand-offs as well as the number of users. Both the variables (number of hand-offs and number of users) are checked to justify that the proposed method is feasible in a large-scale deployment scenario, making the proposed method more robust in practical scenarios. The fluctuations in performance seem common across all of the schemes.

Table 3 summarizes and compares the results of all the parameters, by taking an average for all the problem sizes (such as when user number is 100–500). It can be concluded from the findings discerned from the simulation results that the proposed ACO-based CRE method, along with its closest competitor, the PSO technique, exhibits the lowest call drop rate, the most balanced network when compared to state-of-the-art methods, namely, M-OPTG and M-WO. However, it has been observed that the throughput has dropped for higher number of users in the earlier figures but the average performance

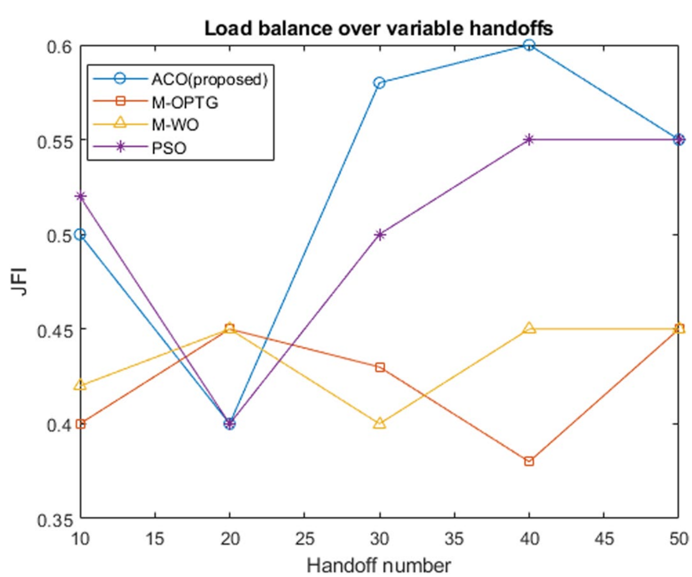


**Fig. 6** Jain's Fairness Index (JFI) over variable hand-off number

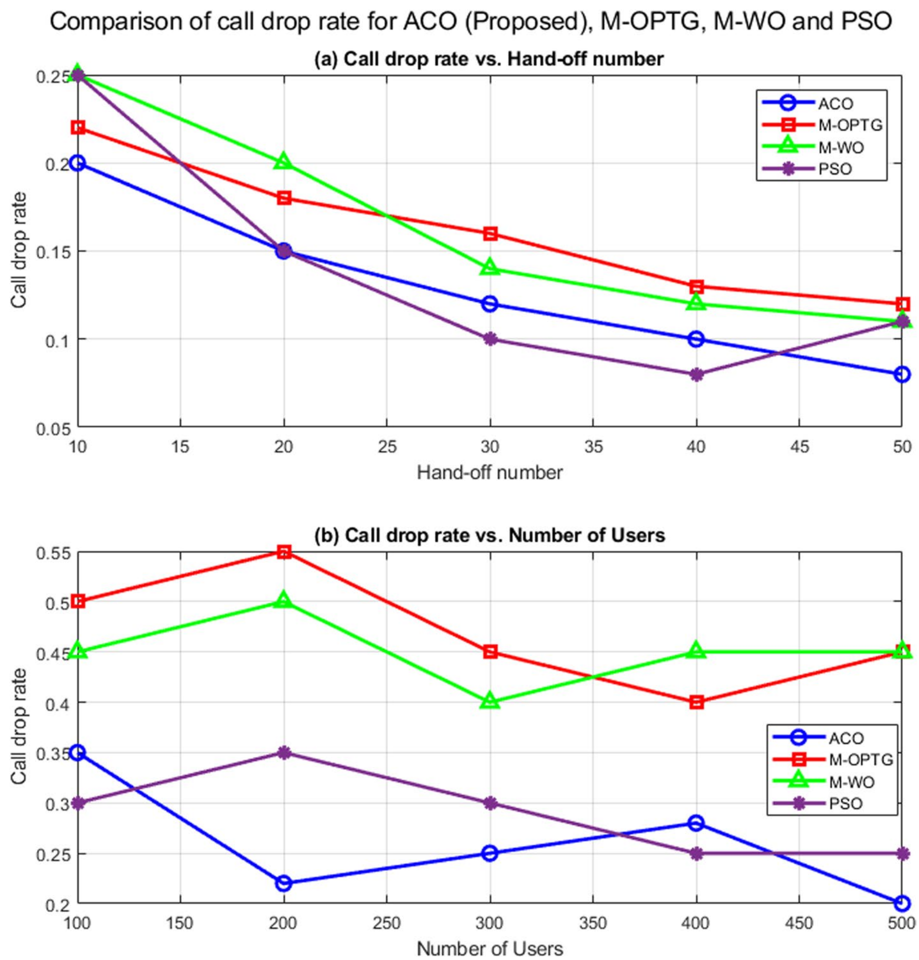


**Fig. 7** Call drop rate evaluation versus **a** hand-off number, **b** number of users for the proposed scheme compared to other recent schemes

**Table 3** Summary of the performance comparison

| Techniques | Throughput (Mbps) | JFI | Call Drop Rate |
|---|---|---|---|
| M-OPTG | 66.2 | 0.614 | 0.47 |
| M-WO | 65.6 | 0.698 | 0.45 |
| PSO | 69 | 0.664 | 0.29 |
| ACO (Proposed) | 70.6 | 0.774 | 0.26 |

with respect to throughput is higher. It can be asserted that, on average, the proposed ACO outperforms all other schemes across all other scenarios considered.

A time complexity comparison is illustrated in Fig. 8. Here, the execution time of the optimization algorithm is verified with respect to the number of users. It can be observed that there is a general trend of increase in execution time with respect to number of users, which is obvious. The rise is initially linear but exponential toward the end for all the algorithms. The algorithms are executed for 100 times and took an average for each instance (such as when the number of users is 100, 200, or 500). The simulations are

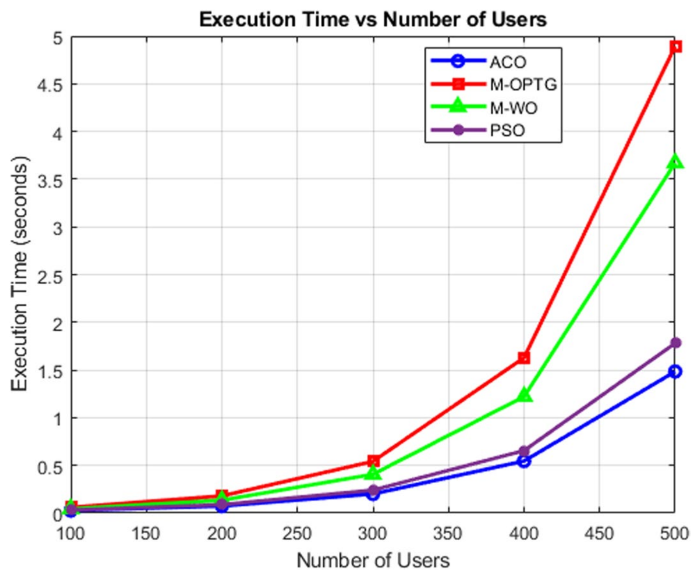


**Fig. 8** Time complexity comparison

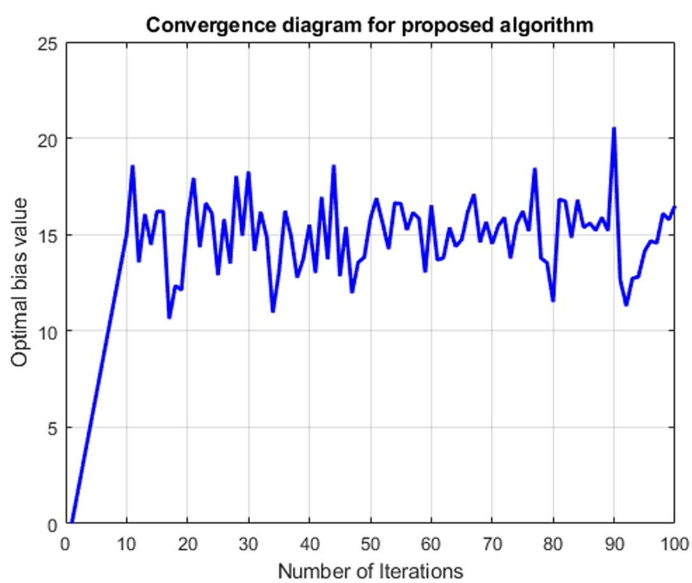


**Fig. 9** Convergence graph for the proposed algorithm

conducted under consistent conditions across all compared algorithms to prevent any ambiguity.

Additionally, a convergence graph for the proposed algorithm is presented in Fig. 9. After the first 10 iterations, the bias values start to fluctuate around the optimal value. This fluctuation is represented by the variations in the blue line around the optimal bias value. The fluctuation occurs due to the inherent randomness of the optimization process as well as the network conditions including user positions. Despite the fluctuations, the bias values eventually stabilize around the optimal value as the number of iterations increases. This stabilization indicates that the optimization algorithm has converged to a solution, and further iterations do not significantly change the bias value. The minimum value of bias for the proposed algorithm found from 100 executions was 10.2, the average value of the bias was 15.3, and the standard deviation was 1.65.

## 7 Conclusion and future work

To conclude, this research has explored the context of the HetNet environment and introduced a novel strategy for load balancing and cell association. Our approach has incorporated a bio-inspired perspective, leveraging the ACO algorithm. In two-tier heterogeneous network scenarios, conventional user association mechanisms are biased toward the highest received signal power, which tends to favor MBSs over SBSs, causing network imbalance. By introducing a user association mechanism that takes into account the number of users satisfying the downlink demands, our proposed ACO-based CRE method effectively redistributes users toward SBSs. This strategy has the potential to rectify network imbalance without resorting to intricate combinatorial optimization problems. Our proposed ACO-based CRE method has exhibited promising outcomes by mitigating network imbalance, while achieving satisfactory throughput levels for individual UEs as well as superior average throughput for the whole network. The call drop rate is also found to be relatively low compared to the state-of-the-art techniques.

Future research could use dynamic clustering methods to simplify complex searches and improve accuracy. Additionally, there can be plans to explore CoMP scenarios involving multiple base stations to user connections to evaluate efficiency with high-speed moving users. As more base stations and users are deployed, future studies might also examine how this approach works in device-to-device and IoT communications. These extensions could optimize user association and help the end users get the enhanced mobile service.

**Abbreviations**

| | |
|---|---|
| QoS | Quality of service |
| HetNet | Heterogeneous network |
| CRE | Cell range extension |
| ACO | Ant colony optimization |
| SINR | Signal to interference-plus-noise ratio |
| IoT | Internet of Things |
| MNO | Mobile network operators |
| BS | Base station |
| UDN | Ultra-dense network |
| 3GPP | 3rd Generation Partnership Project |
| MBS | Macro base station |
| RSRP | Reference signal received power |
| RB | Resource block |
| SON | Self-organized network |
| GA | Genetic algorithms |
| MAC | Media access control |
| KPI | Key performance indicator |
| RFB | Reusable functional blocks |
| SDN | Software designed networks |
| EPC | Evolved packet core |
| PSO | Particle swamp optimization |
| KO | Knapsack optimization |
| M-WO | Multi-objective whale optimization |
| VHO | Vertical hand-off |
| M-OPTG | Multi-objective genetic programming |

**Acknowledgements**

This research project is supported under the Commonwealth Government's Research Training Program. I gratefully acknowledge the financial support provided by the Australian Government and CQUniversity.

**Author contributions**

Mohammed Jaber Alam is the main contributor of this manuscript. He wrote the manuscript, carried out simulations, and undertook analysis of the results. A/Prof Ritesh Chugh revised the manuscript and provided an invaluable guidance in the documentation. Dr Salahuddin Azad and Dr Md Rahat Hossain have reviewed and revised the manuscript as well as validated the results. All authors have read and approved the final manuscript.

**Funding**

Not applicable.

**Availability of data and materials**

Data sharing is not applicable to this article.

## Declarations

**Ethics approval and consent to participate**

Not applicable.

**Competing interests**

The authors have no conflict of interest relevant to this paper.

Received: 10 January 2024 Accepted: 23 May 2024
Published online: 02 June 2024

**References**

1. N. Zohar, Beyond 5g Reducing the handover rate for high mobility communications. J. Commun. Netw. **24**(2), 154–165 (2022). https://doi.org/10.23919/JCN.2022.000001
2. N. Panda, M. Supriya, Efficient data transmission using trusted third party in smart home environments. EURASIP J. Wirel. Commun. Netw. **2022**(1), 118 (2022)
3. E.-K. Hong, I. Lee, B. Shim, Y.-C. Ko, S.-H. Kim, S. Pack, K. Lee, S. Kim, J.-H. Kim, Y. Shin, 6G R &D vision: requirements and candidate technologies. J. Commun. Netw. **24**(2), 232–245 (2022)
4. N. Panda, M. Supriya, Blackhole attack prediction in wireless sensor networks using support vector machine, in *Advances in Signal Processing, Embedded Systems and IoT: Proceedings of Seventh ICMEET-2022* (Springer, 2023), pp. 321–331
5. H. Manzoor, S. Manzoor, N. Ali, M. Sajid, M.I. Menhas, X. Hei, An SDN-based technique for reducing handoff times in WiFi networks. Int. J. Commun. Syst. **34**(16), 4955 (2021)
6. N. Panda, M. Supriya, Blackhole attack impact analysis on low power lossy networks, in *2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT)* (IEEE, 2022), pp. 1–5
7. M.J. Alam, M.R. Hossain, S. Azad, R. Chugh, An overview of LTE/LTE-A heterogeneous networks for 5g and beyond. Trans. Emerg. Telecommun. Technol. **34**, 4806 (2023)
8. Z.H. Abbas, M.S. Haroon, F. Muhammad, G. Abbas, F.Y. Li, Enabling soft frequency reuse and Stienen's cell partition in two-tier heterogeneous networks: cell deployment and coverage analysis. IEEE Trans. Veh. Technol. **70**(1), 613–626 (2020)
9. S. Shibu, V. Saminadan, Ordering parallel successive interference cancellation mechanism to mitigate ICI interference at downlink of the LTE-A HetNet. Int. J. Wirel. Mobile Comput. **24**(3–4), 258–265 (2023)
10. Z.H. Abbas, M.S. Haroon, G. Abbas, F. Muhammad, Sir analysis for non-uniform HetNets with joint decoupled association and interference management. Comput. Commun. **155**, 48–57 (2020)
11. M.J. Alam, A.A. El-Saleh, C.K. Tan, I. Ku, Y.L. Lee, T.C Chuah, Improved joint cell association and interference mitigation for LTE-A heterogeneous networks, in *2018 IEEE 4th International Symposium on Telecommunication Technologies (ISTT)* (2018), pp. 1–4. https://doi.org/10.1109/ISTT.2018.8701715
12. M.J. Alam, M.R. Hossain, R. Chugh, S. Azad, Distance-based cell range extension and almost blank sub-frame for load balancing and interference mitigation in 5g and beyond heterogeneous networks. Eng. Rep. **6**, 12772 (2023)
13. M.J. Alam, A. Gafur, S.Z. Rashid, M. Sadeque, D. Kundu, R. Sayed et al., Permutation based load balancing technique for long term evolution advanced heterogeneous networks. Int. J. Electr. Comput. Eng. **12**(6), 6311 (2022)
14. K. Sheikhidris, M.Y. Alias, M. Roslee, M.J. Alam, Throughput maximization based on user association in heterogeneous networks, in *2018 6th International Conference on Information and Communication Technology (ICoICT)* (2018), pp. 34–37. https://doi.org/10.1109/ICoICT.2018.8528793
15. M.S. Haroon, Z.H. Abbas, G. Abbas, F. Muhammad, Coverage analysis of ultra-dense heterogeneous cellular networks with interference management. Wirel. Netw. **26**(3), 2013–2025 (2020)
16. Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, J.G. Andrews, User association for load balancing in heterogeneous cellular networks. IEEE Trans. Wirel. Commun. **12**(6), 2706–2716 (2013)
17. C. Zhang, P. Patras, H. Haddadi, Deep learning in mobile and wireless networking: a survey. IEEE Commun. Surv. Tutor. **21**, 2224 (2019)
18. Z. Zhang, K. Long, J. Wang, F. Dressler, On swarm intelligence inspired self-organized networking: its bionic mechanisms, designing principles and optimization approaches. IEEE Commun. Surv. Tutor. **16**(1), 513–537 (2013)
19. P. Mahajan, Ant colony optimization for the call drops reduction in wireless heterogeneous networks, in *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)* (IEEE, 2022), pp. 1–6

20. M.J. Alam, M.R. Hossain, R. Chugh, S. Azad, Multi-objective solution of reducing call drop rate and maximizing throughput by ant colony optimization in 5g heterogeneous networks, in *2023 IEEE International Conference on Networking, Sensing and Control (ICNSC)*, vol. 1 (2023), pp. 1–6. https://doi.org/10.1109/ICNSC58704.2023.10319017
21. X. Wang, X. Li, V.C. Leung, Artificial intelligence-based techniques for emerging heterogeneous network: state of the arts, opportunities, and challenges. IEEE Access **3**, 1379–1391 (2015)
22. S. Radouche, C. Leghris, New network selection algorithm based on cosine similarity distance and PSO in heterogeneous wireless networks. J. Comput. Netw. Commun. **2021**, 1–11 (2021)
23. Z. Zhang, W. Huangfu, K. Long, X. Zhang, X. Liu, B. Zhong, On the designing principles and optimization approaches of bio-inspired self-organized network: a survey. Sci. China Inf. Sci. **56**, 1–28 (2013)
24. R. Martínez, A. Mayoral, R. Vilalta, R. Casellas, R. Muñoz, S. Pachnicke, T. Szyrkowiec, A. Autenrieth, Integrated SDN/NFV orchestration for the dynamic deployment of mobile virtual backhaul networks over a multilayer (packet/optical) aggregation infrastructure. J. Opt. Commun. Netw. **9**(2), 135–142 (2017)
25. M. Shojafar, L. Chiaraviglio, N. Blefari-Melazzi, S. Salsano, P5g: a bio-inspired algorithm for the superfluid management of 5g networks, in *GLOBECOM 2017-2017 IEEE Global Communications Conference* (IEEE, 2017), pp. 1–7
26. T.M. Shami, D. Grace, A. Burr, Load balancing and control using particle swarm optimisation in 5g heterogeneous networks, in *2018 European Conference on Networks and Communications (EuCNC)* (IEEE, 2018), pp. 1–9
27. M. Elkourdi, A. Mazin, R.D. Gitlin, Towards low latency in 5g HetNets: a Bayesian cell selection/user association approach, in *2018 IEEE 5G World Forum (5GWF)* (IEEE, 2018), pp. 268–272
28. G. Hattab, D. Cabric, Rate-based cell range expansion for downlink massive MIMO heterogeneous networks. IEEE Wirel. Commun. Lett. **7**(3), 296–299 (2017)
29. W.K. Lai, J.-K. Liu, Cell selection and resource allocation in LTE-advanced heterogeneous networks. IEEE Access **6**, 72978–72991 (2018)
30. K.M. Nasr, K. Moessner, Knapsack optimisation versus cell range expansion for mobility load balancing in dense small cells, in *2018 European Conference on Networks and Communications (EuCNC)* (IEEE, 2018), pp. 1–9
31. T. Velmurugan, S. Khara, S. Nandakumar, D. Sumathi, Seamless vertical handoff using modified weed optimization algorithm for heterogeneous wireless networks. Radioelectron. Commun. Syst. **60**, 431–448 (2017)
32. D. Fooladivanda, C. Rosenberg, Joint resource allocation and user association for heterogeneous wireless cellular networks. IEEE Trans. Wirel. Commun. **12**(1), 248–257 (2012)
33. U. Ghafoor, T. Ashraf, Maximizing throughput and energy efficiency in 6g based on phone user clustering enabled UAV assisted downlink hybrid multiple access HetNet. Telecommun. Syst. **85**, 1–28 (2024)
34. M. Chinipardaz, S. Amraee, A. Sarlak, Joint downlink user association and interference avoidance with a load balancing approach in backhaul-constrained HetNets. PLoS ONE **19**(3), 0298352 (2024)
35. A. Shaverdian, J. Ghimire, C. Rosenberg, Simple and efficient network-aware user association rules for heterogeneous networks. Comput. Netw. **156**, 20–32 (2019)
36. J. Sangiamwong, Y. Saito, N. Miki, T. Abe, S. Nagata, Y. Okumura, Investigation on cell selection methods associated with inter-cell interference coordination in heterogeneous networks for LTE-advanced downlink, in *17th European Wireless 2011-Sustainable Wireless Technologies* (VDE, 2011), pp. 1–6
37. H.P. Kuribayashi, M.A. De Souza, D.D.A. Gomes, K.D.C. Silva, M.S. Da Silva, J.C.W.A. Costa, C.R.L. Francês, Particle swarm-based cell range expansion for heterogeneous mobile networks. IEEE Access **8**, 37021–37034 (2020)
38. A. Oliazadeh, O. Bozorg-Haddad, A. Arefinia, S. Ahmad, Ant colony optimization algorithms: introductory steps to understanding, in *Computational Intelligence for Water and Environmental Sciences* (Springer, 2022), pp. 137–154


## Publisher's Note