Complex Agent-based Modeling for HetNets Design and Optimization

Mostafa Ibrahim*, Umair Sajid Hashmi†, Muhammad Nabeel‡, Ali Imran‡, and Sabit Ekin*
*School of Electrical & Computer Engineering, Oklahoma State University, Oklahoma, USA
†School of Electrical Engineering & Computer Science, National University of Sciences & Technology, Pakistan
‡School of Electrical & Computer Engineering, University of Oklahoma, USA
*{mostafa.ibrahim,sabit.ekin}@okstate.edu,
†umair.hashmi@seecs.edu.pk, ‡{muhmd.nabeel,ali.imran}@ou.edu

2022 1st International Conference on 6G Networking (6GNet) | 978-1-6654-6763-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/6GNet54646.2022.9830485

***Abstract*—In wireless heterogeneous networks (HetNets), complexity is an intrinsic property. This paper presents agent-based modeling (ABM) as a tool to optimize complex HetNets. We introduce and analyze a HetNet ABM model that employs parallel algorithms for interference management, resource allocation, and load balancing at both micro and macro levels. Two reinforcement learning (RL) algorithms jointly work together in the model to resolve co-tier and cross-tier interferences. The first RL algorithm controls the transmission power of the small cells, whereas the second assigns the users to the sub-bands with less interference levels. Concurrently, the user association is decided by the users based on their preferences and the resources available at the cells. The model is analyzed in three different operation modes, by switching processes on and off. Results show that individual processes contribute to overall system performance, while jointly maximizing the network's aggregate signal-to-interference-and-noise ratio (SINR) and minimizing load-induced latency by efficient load balancing.**

I. Introduction

Heterogeneous networks (HetNets) and small cell densification are the key components of 5G and Beyond wireless networks. The goal of cell densification is to improve network parameters including capacity, coverage, latency, and load distribution. However, a number of technological challenges constrain the deployment of small cell networks. The two most critical challenges discussed in the literature include interference management and self-organization [1], [2]. The self-organization, self-configuration, and self-analysis capabilities are important as they significantly contribute toward the overall network performance.

While optimizing a HetNet, there is a certain criteria to consider and several trade-offs to resolve [3]. The goal is to jointly optimize different HetNet parameters for better interference management [4], user quality of experience [5], resource allocation [6], latency [7], user association [8], cell load balancing [9], energy efficiency [10], [11], mobility and handovers [12], costs of deployment [13], optimal efficiency trade-offs [14], and coexistence with other radio access technologies [15]. Therefore, a suitable modeling framework is needed to formalize the multi-dimensional optimization problem completely and then solve it to yield optimal operating parameters.

In the literature, many simulation paradigms have been presented for such dynamic cases [16]. The game-theoretic system is one of the major modeling paradigms [17] that study strategies and interactions among players who behave rationally in order to maximize their benefits [18]. The assumption of purely rational agents, on the other hand, is not necessarily represented in practical networks.

Multi-agent Reinforcement Learning (RL) is a common machine learning-based paradigm for HetNets. This paradigm depends on players making decisions in their environment to maximize their utility function [19]. A multi-agent RL framework faces several challenges [20], [21]. For the 5G and beyond HetNets, high-dimensional state and action space adds non-practical computational complexity and long learning time. Another difficulty is choosing the reward functions, particularly when we have several types of agents.

In this paper, we exploit agent-based modeling (ABM) to address the aforementioned optimization problem. ABM is a tool that studies a complex system's emergent activity on a macro level by modeling micro-scale interactions within a population of agents [22]. ABMs are studied in simulation environments, with players/agents following laws that do not necessarily relate to utility functions [23]. Unlike game theory, ABMs allow the designer to model different interacting games within the same model without creating an analytical framework. It also allows testing of various player heuristics without assuming cognitive abilities. Therefore, it can implement real industry scenarios and evaluate them across all the network parameters.

The main contributions of this work can be summarized as follows:

- We propose a modeling paradigm that considers the intrinsic complexity of HetNets. It has the capability to incorporate a diversity of game-theoretic, machine learning, and rule-based algorithms within the same model. Which was not possible before with analytical models.
- We then develop a novel agent-based modeling (ABM) approach to examine and analyze the complex interactions of HetNet nodes. The network nodes in this model are running in parallel as independent entities and the learning algorithms are running concurrently.

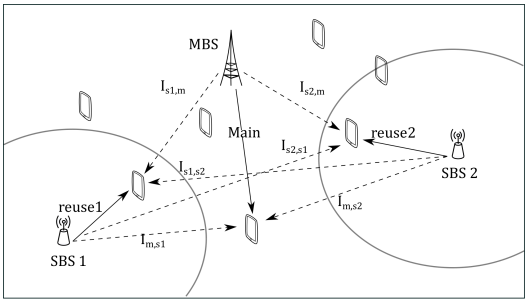


Fig. 1: Two tier network architecture, representing the main (desired) link as a solid line and interferers with dotted lines.

- We simulate the model using a stochastic geometry based environment. Our results show that the proposed approach using two concurrent reinforcement learning based algorithms offers efficient resource allocation and maximizes the network's aggregate signal-to-interference-and-noise ratio (SINR).

The rest of the paper's organization is as follows. In Section II, the HetNet system model is presented. The proposed agent-based model is discussed in Section III, whereas Section IV is dedicated to simulations and results. Finally, we conclude the paper in Section V.

II. System Model

In this section, a distributed system model is proposed to study a complex practical HetNet. The two-tier system is composed of three types of agents: macrocells (MCs), small cells (SCs), and user equipments (UEs). The spectrum is shared between macrocells and small cells, and is reused several times within the same macrocell to increase the network spectral efficiency.

A. Network Model

The modeled HetNet is a 2-tier network with macrocells forming the main network and small cells used as the second tier cells, as shown in Fig. 1. A macrocell's assigned spectrum is reused at the lower teir. The system is based on the long-term evolution (LTE) time-frequency resource block numerology. The full network spectrum is used orthogonally between macrocells.

B. Cell Association

UEs have different preferences regarding SINR, latency, and the number of requested resource blocks (RBs). Affected by what the cells are offering, the UEs decide the cell association. The cells have the responsibility of coordinating and distributing the spectrum between each other. Also, they manage the network load balancing and interference levels at the UEs.

C. Channel Model

The large scale path loss PL used in our model is the simplified free space model: $PL(dB) = \kappa + 10\zeta \log_{10}(d)$, where d is the distance between the UE and the serving cell, ζ is the path loss exponent, and κ is a unitless factor that depends on the average channel attenuation, frequency of operation, and antenna characteristics.

In the presented downlink scheme, the interferences induced by spectrum reuse are cross-tier interference and co-tier interference. For cross-tier interference, at small cell s_i user from macrocell m is given as $I_{s_i,m}$. The interference from a small cell to a macrocell user is given as I_{m,s_i}. In comparison, the co-tier interference from a small cell to a user of another small cell is given as I_{s_i,s_j}. The main (desired) link is represented as a solid line in Fig. 1, while the interference links are represented with dotted lines. The SINRs $\gamma_{n,m}$, and $\gamma_{n,s}$ at the nth user served by macrocell m and the small cell s, on the rth resource block, are formalized respectively as:

$$\gamma_{n,m}^{(r)} = \frac{|h_{n,m}^{(r)}|^2 p_m^{(r)}}{N_{n,m}^{(r)} + \sum_{s\in S} |h_{n,s}^{(r)}|^2 p_s^{(r)}} \ , \tag{1}$$

$$\gamma_{n,s}^{(r)} = \frac{|h_{n,s}^{(r)}|^2 p_s^{(r)}}{N_{n,s}^{(r)} + \sum_{m\in M} |h_{n,m}^{(r)}|^2 p_m^{(r)} + \sum_{j\in S, j\neq s} |h_{n,s}^{(r)}|^2 p_s^{(r)}} \ , \tag{2}$$

where S is the set of small cells and M is the set of macrocells, $N_n^{(r)}$ is the noise variance, and $h_{i,m}$, and $h_{i,s}$ are the channel coefficients from the macrocell and small cells, respectively, to user n. $p_s^{(r)}$ are the transmit powers of the macrocell and small cells over resource block r, respectively.

D. User Requests

The user n creates u requests per unit time t, with rate λ_r. This random variable u follows a Poisson process. For each user, the number of requested resource blocks x is a truncated normal distribution over the interval $0 < x < \infty$, with mean $\mu_x^{(n)}$ and standard deviation $\sigma_x^{(n)}$.

III. Proposed Agent Based Architecture

The proposed system architecture is described with several processes performed by each agent (UE, MC, or SC), and a set of interactions between those agents. Each process is formalized with a flowchart. Hence, agent's behavior can be summed by several processes running asynchronously and in parallel.

A. User Terminals ABM Process

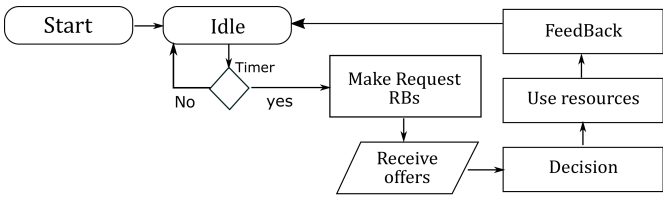


Fig. 2: User equipment flowchart: Service request and usage.

A UE is assumed to be exchanging information with several nearby cells (macrocells or small cells) over the control channel, but it receives the service only from one of them.

The flow chart in Fig. 2 describes a resource block request and utilization cycle. The UE is initially at the *'Idle'* state. Then it moves to the *'Make Request RBs'* state when a countdown timer reaches zero. The timer value is a random variable τ that corresponds to the interval time between two requests. It is assigned a new value after the timer expires, following the exponential distribution: $f(\tau) = \lambda_r^{(n)} e^{-\lambda_r^{(n)}\tau}$, where $\lambda_r^{(n)}$ is the request rate for user n. This ensures a Poisson distribution for the number of requests per unit time.

The cells then send offers that depend on their transmit power levels, as we will observe in the following sections. The offer is composed of the number of offered resource blocks RBs, the start and end in the frequency domain given by f_1 and f_{end} respectively, the start and end in the time domain given by t_1 and t_{end} respectively, and the cell transmit power P_{tx}. The UE then collects all the offers at the *'Receive offers'* state, chooses the best offer, and sends an accept response to the corresponding cell. The UE chooses the best offer based on the following mathematical utility function:

$$U(c) = \arg\max_{c} \left(\frac{RBs_c \times w_r \log_2(1 + 10^{(\gamma_c/10)})}{1 + w_d(t_{end} - t)} \right), \quad (3)$$

where the subscript c corresponds to cells. This function sets the UE service preferences by assigning the weights: w_r for the expected throughput at the receiver, and w_d for latency. The w_r and w_d values are proportional to the importance of each corresponding factor to the UE. Note that the value $(t_{end} - t)$ represents how long it takes for the RBs to reach the UE. During the *'Use resources'* state, the UE measures the quality of service affected by the interference levels. Then it is shared with the serving cell in the *'Feedback'* state. The UE reports its feedback to the serving cell before returning to *'idle'*. The feedback holds information about the interference levels, the SINR, and the delay.

B. Sub-band Management at the Macro- and Small Cells

In our design, the cells are responsible for two main tasks: interference management and cell load balancing. In a spectrally efficient system, the small cells share the spectrum with the upper-tier (macrocells). In the downlink scheme, the macrocells' bands are divided into sub-bands SB higher in granularity than a resource block. The sub-bands are reused for several times. For a single reuse case, the small cells and the macrocells coordinate to minimize the cross-tier interference by adjusting the small cells' transmit power levels. The macrocell users' feedback on the interference levels is used to adjust the small cells' transmission powers. The decisions for power level allocations are taken at the MCs level. For a second reuse case, the sub-band will be used twice at two different small cells. More reuse levels increases the interference levels between the network cells, and enlarges the optimization space. Therefore, the twice reuse case is the one evaluated in the simulations section.

C. Reinforcement Learning Processes

Reinforcement learning (RL) is used for two processes; first, to adjust the power levels in the reuse schemes; second, to assign the users to the sub-bands with highest performance level. We represent those RL processes in the following two subsections.

1) Small Cell Transmit Power Management : This process is running under the macrocell agents. Initial power levels are assigned for the small cells. Then it enters a loop of collecting rewards and updating the small cell power values. Due to the nature of the problem, the multi-armed bandit method is used as our model-free reinforcement learning method [24]. A multi-armed bandit algorithm has a number of actions to choose from, hence the term 'arm'. Learning is done over rounds; in each round, depending on the exploration factor ϵ, an arm is chosen, and the corresponding reward R_i is collected during the round duration.

The macrocell's multi-armed bandit algorithm list of actions is $A_i = \{a_i^{(p)}\}_{p \in \{P_1, P_2, \ldots, P_k\}}$, where $a_i^{(p)}$ represents the power transmit level for the reused ith sub-band SB_i, from a set of transmit power levels, and k is the number of the power levels. The value function Q holds an evaluation for the expected reward for each action.

The value function is updated via the recursive equation:

$$Q_{t+1}(A_i) = (1 - \alpha) Q_t(A_i) + \alpha(R_i) \ , \quad (4)$$

where α is a learning-discount factor.

The reward function used for the proposed model is the aggregate SINR for all RBs in sub-band SB_i, over the last learning episode T_e: $R_i = \sum_{t_1 > t - T_e} \sum_{RB \in SB_i} \gamma_{RB,t_1}$.

The power management RL algorithm is shown below in Algorithm 1. Deploying this algorithm determines the proper reuse power levels to achieve the maximum reward over each sub-band.

Algorithm 1 Small Cell Sub-band Power Management

```
Initialization Q(A_i) = 0 ∀ i
Initialize the reuse power levels P(SB)
For each Sub-band SB define power levels list A_i
while True do
    for i ∈ Sub-bands do
        if rand(.)< ε then
            Explore: choose action from A_i randomly
        else
            Exploit: choose action A_i(t+1) = arg max_{a_i} Q_{t+1}
        end
        Receive rewards R_i(t + 1), and update Q table
    end
end
```

In the second reuse case, two small cells transmit different power values for each sub-band. Therefore, the action space is two dimensional: $a_{i,j} \in$

Algorithm 2 User sub-band choice learning algorithm.

initialization $Q(A_n) = 0 \ \forall \ n$
define list of sub-bands at this cell
while *True* **do**
  if *rand(.)*< ϵ **then**
    Explore: Chose action from A_n randomly
  else
    Exploit: Choose action $A_n(t+1) = \arg\max_{a_n} Q_{t+1}$
  end
  Receive rewards $R_n(t+1)$, and update Q table
end

$[(P_{S_{i_1}}, P_{S_{j_1}}), (P_{S_{i_1}}, P_{S_{j_2}}) \ ... \ (P_{S_{i_K}}, P_{S_{j_K}})]$, where i and j are the notation of the same sub-band for two different small cells, S_i and S_j.

2) User to Sub-band Association: UEs can have different performance levels for different sub-bands at the same cell. This is affected by the distribution of the set of users served by the cell and their distances from the interfering cell. Therefore, this method is proposed to allocate each UE on the sub-band that suits its position with respect to the other agents (cells and UEs) in the network.

A multi-armed bandit process starts by assigning the served users to the available sub-bands randomly. Then it keeps collecting the service feedback from the UEs.

The rewarding functions are formulated from the collected feedback. Each UE has its own Q-table that gets updated from the reward functions. The Q-table holds the values reflecting the learned performance per sub-band. The learning algorithm components for the nth user include actions $A_n = \{a_n^{(s)}\}_{s\in\{1,...,N_S\}}$, where $a_n^{(s)}$ represents the action of switching to one of the cell sub-bands, rewards R_n, value function Q, and explorer factor ϵ.

The proposed reward value for this algorithm is: $R_n = \frac{\gamma_n}{1 + w_{d_n} t_{d_n}}$, where γ is the SINR, and t_d represents the delay experienced by the UE during the last served RBs. The factor $(1 + w_d t_d)$ normalizes the SINR level by the latency level to ensure that the users associate to the sub-bands, not only based on the SINR but also the sub-band load induced latency. The learning algorithm is described in Algorithm 2.

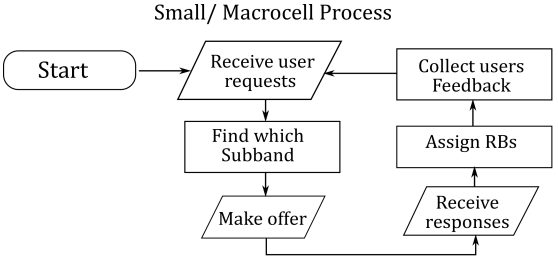


Fig. 3: RB assignment flow chart.

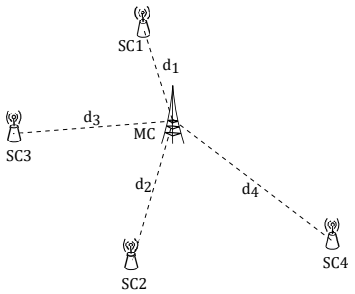


Fig. 4: Simulation environment.

D. Resource Blocks Allocation ABM Process

We close the system model with the small cell RB assignment process illustrated in Fig. 3, which basically elaborated on the mechanism of the SC response to UE requests. The SC process keeps listening to the UE requests. Once it receives a request, it finds the sub-band which is suitable for this UE. The suitable sub-band is already determined in the RL process described in Section III-C2.

Now based on the SC current load, an offer is formulated. Ideally, if the SC is not congested, the offered RBs will be the same number as the requested RBs. However, if the SC is congested (cell load > specific value L_h), a discounted number $((1 - D_{factor}) \times RBs)$ is offered.

Finally, after the UE has finished using the RBs, the SC receives the users' feedback.

IV. Results and Discussion

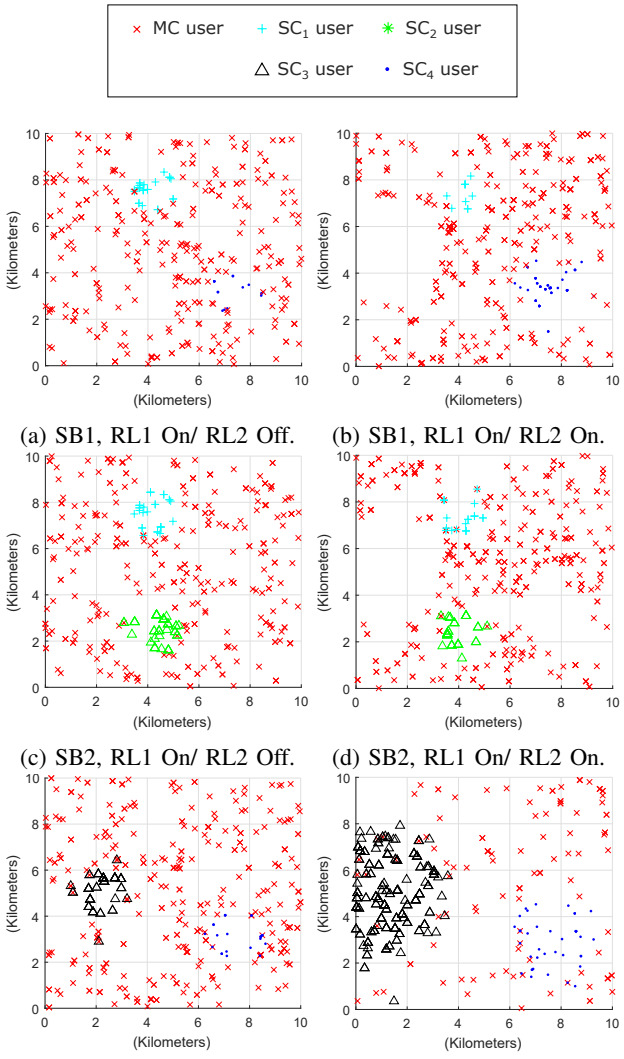
Parameters	Values
Users positions	uniformly distributed in the area [x=[0, 10] km, y=[0, 10] km]
Number of users	500 users
Macrocell tx power	30 dBm
Small cell tx power range	[15, 25] dBm
Pathloss model	$25 \log_{10}(d) + 40$
Number of RBs per sub-band	10 RBs
Request rate for UEs	$\lambda_r = 0.012$ request per T
T	RB duration
Requested RBs statistics	Avg. RBs per request= 4 , Std. dev. RBs per request= 2
RL1 Learning episode length	200 T
RL1 explore factor (ϵ)	decreasing from 1 to 0 over [0 7000 T]
RL2 Learning episode length	1 UE-request cycle
RL2 explore factor (ϵ)	fixed 0.3
RL1 and RL2 α factors	$\alpha_1 = \alpha_2 = 0.3$

TABLE I: Simulation parameters.

In this section, we developed the proposed ABM architecture shown in Fig. 4 and simulated it with the parameters in Table 1. We have a MC in the middle and four SCs with distances $d_1 = 2.6 \ km, d_2 = 2.7 \ km, d_3 = 2.8 \ km$, and $d_4 = 2.9 \ km$. An environment module is responsible for instantiating the agent instances and managing the order of calling those objects. Parallelism is emulated by discretizing the time into units, and the environment loops over all the agent instances in each time unit (also called tick). The

network has three sub-bands that are reused twice between the macrocell and the small cells.

The system is evaluated under two different modes of operation. In the first mode, the first learning algorithm (RL1), responsible for power management, is enabled, and the second algorithm (RL2), responsible for user sub-band association, is disabled. In the second mode, both the algorithms RL1 and RL2 are enabled.



(a) SB1, RL1 On/ RL2 Off. (b) SB1, RL1 On/ RL2 On.

(c) SB2, RL1 On/ RL2 Off. (d) SB2, RL1 On/ RL2 On.

(e) SB3, RL1 On/ RL2 Off. (f) SB3, RL1 On/ RL2 On.

Fig. 5: Users distribution for three sub-bands.

The resultant user distribution between the three sub-bands is shown in Fig. 5, where a comparison between the two modes is demonstrated side-by-side in geographical space. For the first mode (Fig. 5a, 5c, and 5e), the macro cell UEs geographical distribution is more uniform than in the second mode. In the second mode (Fig. 5b, 5d, and 5f), due to the second RL algorithm, the macrocell UEs avoid the small cell interference, and they move to sub-bands with less interference. Which allows the small cells to transmit with higher power levels resulting in more coverage, hence more users. Next, we plot the complementary cumulative distribution function (CCDF) for the latency experienced by the UEs in Fig. 6. Latency is induced by request queuing at the high-loaded cells. We see that in the case of the second mode, UEs experience less latency due to better load balancing. Then

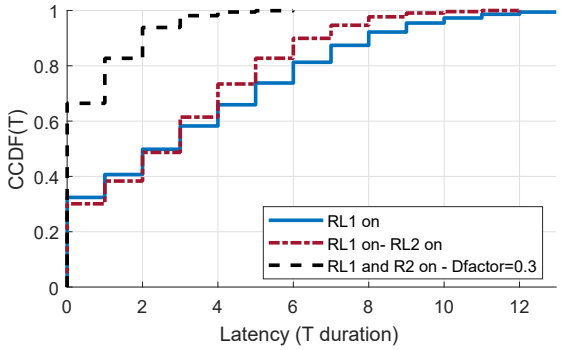


Fig. 6: Latency CCDF.

as an attempt to have better load balancing, we increase the discount factor to 0.3 for a load above $L_h = 50\%$ for all the cells. The resultant latency CCDF is shown in Fig. 6, whereas the corresponding aggregate SINR, average latency, and tier loads are listed in the summarizing Table II. We observe that this operation mode has lower latency and the highest SINR and load balancing, on the expense of discounted number of RBs served by the whole network.

TABLE II: Results Summary

RL1	RL2	D-factor	Aggregate SINR (dB)	SCs load	MC load	Average latency
ON	OFF	0	57 dB	16 %	90 %	3.3 T
ON	ON	0	62 dB	23 %	74%	2.8 T
ON	ON	0.3	66 dB	21%	67%	0.6 T

Below, we compare with similar systems proposed in the literature. These frameworks have been adapted in our architecture to be comparable with our proposed RL algorithms. For the first study [25], the utility function of Algorithm 1 is replaced with their proposed utility function:

$$U_1 = \arg\max_{p_i \in P} \sum_{t_1 > t - T_e} \sum_{k \in K} \log_2(1 + \gamma_k^{(RB)}) \mathbb{1}_{\{\gamma_m^{(RB)} > \Gamma_{th}\}}, \tag{5}$$

Algorithm 2 is deactivated as it has no relevance to this study. The second framework performs inter-cell interference coordination ICIC, [26], [27]. Like our study, it is composed of two parts: sub-channel allocation and power assignment algorithms. The utility function used for Algorithm 1 and 2 are as follows:

$$V_1 = \arg\min_{p_i \in P} \sum_{t_1 > t - T_e} \sum_{RB \in SB_i} I_{RB,t_1} + (w_{d_{ICIC}} t_d) \ , \tag{6}$$

$$V_2 = \arg\min_{s_n \in S} \left(I_{RB} + (w_{d_{ICIC}} t_d)\right) \ . \tag{7}$$

A delay factor $\omega_{d_{ICIC}}$ was added to manage the latency induced by unbalanced load distribution.

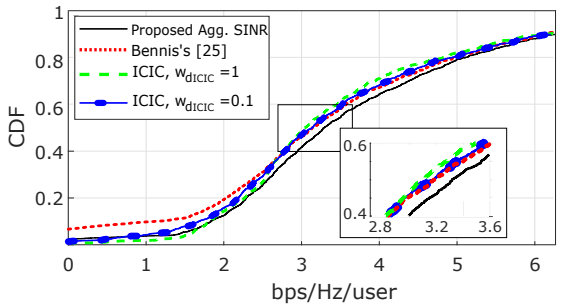


Fig. 7: Comparison with existing literature: Per-user throughput CDF.

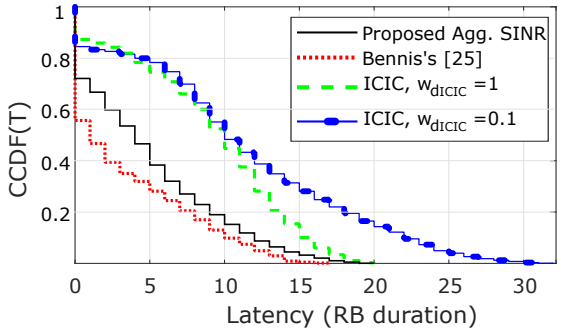


Fig. 8: Comparison with existing literature: latency CCDF.

The results of the comparisons are shown in Fig. 7 and Fig. 8, for the per-user throughput CDF, and per-user latency CCDF. The average throughput is found to be higher for the maximum aggregate SINR utility function case. On the other hand, the ICIC framework has a slightly lower number of low throughput users. This is due to the focus of the ICIC algorithm on minimizing the interference. The work in [25] has a higher number of low throughput users due to not deploying a sub-band or a sub-channel algorithm, as in Algorithm 2. The per-user latency CCDF is a measure of the efficiency of load distribution between the macrocell and small cells; and amongst small cells. The two aggregate SINR based methods achieved lower latency values than the minimum interference-based method. The usage of Algorithm 2 added latency due to the users of preferring to utilize sub-bands that are not reused more than the reused sub-bands. This results non-uniformity in sub-band utilization, hence the slight increase in latency. In the case of the ICIC algorithm this imbalance can be managed by modifying the utility function in Algorithm 2 to take sub-band association decisions based on the delay. Hence, we can also observe the effect of w_d on the latency results.

V. Conclusion

This paper sheds light on the complex nature of HetNets and proposes an ABM framework through which a complex dynamic network can be formalized. Agent-based modeling is a computational method that can create extensive models with various levels of rationality at the agents. It incorporates rule-based behaviors and learning algorithms within the same model. We also proposed a client-driven system model, in which cells control power and spectrum based on user requests and feedback. The proposed ABM uses two concurrent reinforcement learning based algorithms offering efficient resource allocation, interference management, and load balancing. The first RL algorithm on a multi-armed bandit problem was used to manage the transmit powers of small cells in order to maximize the network's aggregate throughput. The other RL algorithm was used to drive user sub-band association in order to maximize SINR while minimizing user latency. In the simulations section, the emergent behavior was shown in the users' distribution within sub-bands and geographical space. Also, The coordination gain between the two learning algorithms was shown. Further, we show that the discounted offer rule's adds to the network aggregate SINR, and enhances load balancing, and load induced latency performance. Finally, a comparison with similar work in the literature was performed for more insight on the enhancements of our work.

Acknowledgment

The work was supported by the National Science Foundation under Grants 1923295 and 1923669.

References

[1] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, Present, and Future," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 497–508, 2012.

[2] D. Lopez-Perez, A. Valcarce, G. de la Roche, and J. Zhang, "OFDMA femtocells: A roadmap on interference avoidance," *IEEE Communications Magazine*, vol. 47, no. 9, pp. 41–48, 2009.

[3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What Will 5G Be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[4] N. Ksairi, P. Bianchi, P. Ciblat, and W. Hachem, "Resource Allocation for Downlink Cellular OFDMA Systems—Part I: Optimal Allocation," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 720–734, 2010.

[5] U. S. Hashmi, A. Rudrapatna, Z. Zhao, M. Rozwadowski, J. Kang, R. Wuppalapati, and A. Imran, "Towards Real-Time User QoE Assessment via Machine Learning on LTE Network Data," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, 2019, pp. 1–7.

[6] D. Kivanc, Guoqing Li, and Hui Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Transactions on Wireless Communications*, vol. 2, no. 6, pp. 1150–1158, 2003.

[7] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and Low-Latency Wireless Communication: Tail, Risk, and Scale," *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, 2018.

[8] S. Cetinkaya, U. S. Hashmi, and A. Imran, "What user-cell association algorithms will perform best in mmWave massive MIMO ultra-dense HetNets?" in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–7.

[9] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. S. Dhillon, "An overview of load balancing in hetnets: old myths and open problems," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 18–25, 2014.

[10] Y. Liu, C. S. Chen, C. W. Sung, and C. Singh, "A Game Theoretic Distributed Algorithm for FeICIC Optimization in LTE-A HetNets," *IEEE/ACM Transactions on Networking*, vol. 25, no. 6, pp. 3500–3513, 2017.

[11] F. Albiero, F. H. P. Fitzek, and M. Katz, "Cooperative Power Saving Strategies in Wireless Networks: an Agent-based Model," in *2007 4th International Symposium on Wireless Communication Systems*, 2007, pp. 287–291.

[12] M. Tayyab, X. Gelabert, and R. Jäntti, "A Survey on Handover Management: From LTE to NR," *IEEE Access*, vol. 7, pp. 118 907–118 930, 2019.

[13] J. Zausinova, M. Zoricak, V. Gazda, G. Bugar, and J. Gazda, "An Agent-Based Model of Adaptive Pricing in HetNets," in *2019 3rd International Conference on Advanced Information and Communications Technologies (AICT)*, 2019, pp. 31–35.
[14] U. S. Hashmi, S. A. R. Zaidi, and A. Imran, "User-Centric Cloud RAN: An Analytical Framework for Optimizing Area Spectral and Energy Efficiency," *IEEE Access*, vol. 6, pp. 19 859–19 875, 2018.
[15] M. Yan, G. Feng, and S. Qin, "Multi-RAT Access Based on Multi-Agent Reinforcement Learning," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–6.
[16] S. E. Page, "Uncertainty, difficulty, and complexity," *Journal of Theoretical Politics*, vol. 20, no. 2, pp. 115–149, 2008.
[17] O. Morgenstern and J. Von Neumann, *Theory of games and economic behavior*. Princeton university press, 1953.
[18] U. S. Hashmi, A. Islam, K. M. Nasr, and A. Imran, "Towards User QoE-Centric Elastic Cellular Networks: A Game Theoretic Framework for Optimizing Throughput and Energy Efficiency," in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2018, pp. 1–7.
[19] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.
[20] T. T. Nguyen, N. D. Nguyen, and S. Nahavandi, "Deep Reinforcement Learning for Multiagent Systems: A Review of Challenges, Solutions, and Applications," *IEEE Transactions on Cybernetics*, pp. 1–14, 2020.
[21] G. Dulac-Arnold, D. Mankowitz, and T. Hester, "Challenges of real-world reinforcement learning," *arXiv preprint arXiv:1904.12901*, 2019.
[22] M. Ibrahim, U. S. Hashmi, M. Nabeel, A. Imran, and S. Ekin, "Embracing complexity: Agent-based modeling for hetnets design and optimization via concurrent reinforcement learning algorithms," *IEEE Transactions on Network and Service Management*, vol. 18, no. 4, pp. 4042–4062, 2021.
[23] J. H. Holland, *Complexity: A very short introduction*. OUP Oxford, 2014.
[24] T. Lattimore and C. Szepesvári, *Bandit algorithms*. Cambridge University Press, 2020.
[25] M. Bennis, S. M. Perlaza, P. Blasco, Z. Han, and H. V. Poor, "Self-Organization in Small Cell Networks: A Reinforcement Learning Approach," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3202–3212, 2013.
[26] H. Zhang, Y. Wang, and H. Ji, "Resource Optimization-Based Interference Management for Hybrid Self-Organized Small-Cell Network," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 2, pp. 936–946, 2016.
[27] M. Simsek, M. Bennis, and I. Guvenc, "Learning Based Frequency- and Time-Domain Inter-Cell Interference Coordination in HetNets," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4589–4602, 2015.