IEEE 5G World Forum
July 9-11, 2018

Towards Low Latency in 5G HetNets: A Bayesian Cell Selection / User Association Approach

Mohamed Elkourdi, Asim Mazin, and Richard D. Gitlin, *Life Fellow, IEEE*
Innovation in Wireless Information Networking Lab (*i*WINLAB)
Department of Electrical Engineering, University of South Florida, Tampa, Florida 33620, USA
Email:{elkourdi, asimmazin}@mail.usf.edu, richgitlin@usf.edu

***Abstract*—Expanding the cellular ecosystem to support an immense number of connected devices and creating a platform that accommodates a wide range of emerging services of different traffic types and Quality of Service (QoS) metrics are among the 5G's headline features. One of the key 5G performance metrics is ultra-low latency to enable new delay-sensitive use cases. Some network architectural amendments are proposed to achieve the 5G ultra-low latency objective. With these paradigm shifts in system architecture, it is of cardinal importance to rethink the cell selection / user association process to achieve substantial improvement in system performance over conventional maximum signal-to- interference plus noise ratio (Max-SINR) and cell range expansion (CRE) algorithms employed in Long Term Evolution-Advanced (LTE-Advanced). In this paper, a novel Bayesian cell selection / user association algorithm, incorporating the access nodes capabilities and the user equipment (UE) traffic type, is proposed in order to maximize the probability of proper association and consequently enhance the system performance in terms of achieved latency. Simulation results show that Bayesian game approach attains the 5G low end-to-end latency target with a probability exceeding 80%.**

I. Introduction

Fifth-generation (5G) networks are expected to support a broad domain of emerging services with various Quality of Service (QoS) requirements, i.e., from narrow-bandwidth, delay-sensitive services to bandwidth-hungry, delay-tolerant services. To address these disparate services, several solutions have recently been proposed. For instance, Coordinated Multi-Point (CoMP) [1] is envisioned as one of the prominent 5G solutions by improving the service for cell edge users suffering from high levels of interference particularly in multi-tier networks. CoMP has been standardized in LTE-Advanced (Release.11) [2].

Moreover, 5G is anticipated to witness an increase in the heterogeneity and density of access nodes (ANs) (e.g. ultra-dense networks (UDNs)) [3] as a solution to cope with the tremendous growth in the number of devices connected to the network and consequently boost the system's capacity [4].

However, in UDNs the multi-tier interference becomes more severe. Hence, the radio resource allocation process should be carried out in a central unit (CU), while taking into account the bigger network picture [5]. For instance, [6] proposes exploiting the temporal and spatial traffic fluctuations in the network to reduce the interference levels by turning off some ANs with low or no traffic loads.

Recently, the Cloud-Radio Access Network (C-RAN) architecture has attracted attention as a key enabler in implementing interfernce avoicdence and cancellation algortihem for flexiable multi-tier 5G networks [7]. Despite the C-RAN's centralization throughput gain, which is achieved by pooling the computational resources, there is a challenge regarding the delay-sensitive traffic due to the large network latency accruing from transferring the data through the core network and the Internet backbone. To satisfy the low latency requirements of delay-sensitive traffic, some computing processes, application servers and caching capabilities can be migrated from the cloud to the edge of the network as an evolved architecture [8]. This architecture is commonly referred to as a Fog-Radio Access Network (F-RAN) [9], [10]. Hence, some of the traffic, in particular delay-sensitive traffic, can get served at the F-RAN nodes and do not need to travel to the core network and the Internet. This results in a significant reduction in the network latency [11].

As a natural result of the expected increase of heterogeneity in 5G networks, user equipment (UE) will have a large range of connectivity options with different characteristics such as power consumption, latency budget, and the achievable data rate. To best exploit the available opportunities, both the UE's traffic type and the characteristics of available ANs should be taken into account at the cell selection / user association stage.

In traditional cellular systems, the process of cell selection / user association is based on the AN that can provide the highest signal-to-interference-plus-noise ratio (SINR) [12]. However, this approach is generally not optimum in multi-tier networks with diverse traffic. To alleviate this problem, [13] introduces the cell range expansion (CRE) approach for Low Power Nodes (LPNs) via a biasing method such that the High Power Node (HPN) transmit power is reduced on a group of sub-carriers in order to enable better coverage on the same group of sub-carriers for an overlaid LPN. Another approach for cell association is to employ user-perceived rate considering the SINR and the network load [14].

However, these papers did not take into account the different types of traffic. Indeed, the selection and association procedure based on the maximum SINR and CRE criteria only might degrade the performance of the system from a latency perspective (i.e. when UEs of delay-tolerant traffic get associated with Fog-Low Power Nodes (F-LPNs) or when UEs of delay-sensitive traffic are associated with a HPN).

The main contributions of this paper are threefold. First, a system model that supports diverse traffic types is presented. Following that a novel cell selection / user association algorithm based on Bayesian game from the domain of Game Theory is proposed, while incorporating the the traffic type and ANs capabilities. Finally the superiority of the proposed Bayesian cell selection / user association algorithm is proven interms of the achieved latency and the probability of proper assocaition for delay-sensitive and delay-tolerant traffic.

The reminder of this paper is organized as follows. Section II describes the system model. Section III defines the problem and the proposed Bayesian Game approach. In this part, the utility functions of the UE and the network are defined. Section IV investigates the performance of the proposed Bayesian game algorithm through simulations. Finally, the contributions of the paper are summarized in section V.

II. System Model

We consider a simplified two-tier heterogeneous network (HetNet) consisting of one HPN overlaid by several F-LPNs with caching and computation capabilities as shown in Fig.1. The set of all radio access nodes (ANs) in the HetNet is defined as $\mathcal{N} = \{n_0, n_1, ..., n_N\}$, where n_0 represents the HPN and the subset $\mathcal{L} = \{n_1, ..., n_N\}$ denotes the F-LPNs. The F-LPNs are randomly distributed in the service area with spatial density of λ (F-LPNs/km^2). The set of all UEs under the coverage area of the two-tier HetNet is denoted by $\mathcal{U} = \{u_1, ..., u_K\}$. Each UE $k \in \mathcal{U}$ requests a service class defined as the tuple $\phi_k = (\eta_k, \tau_k)$, where η_k, τ_k are the UE k required data rate and latency respectively. Hence, the UE's traffic in the proposed system model, can be mainly classified as delay-sensitive (DS) or delay-tolerant (DT) according to their latency requirement τ.

The signal to interference plus noise ratio (SINR) at UE k, associated with access node $n \in \mathcal{N}$ whose transmitting in downlink on the resoruce block r, is expressed as

$$\gamma_{nk}^r = \frac{h_{nk} P_{nk}^r}{\sum_{n \in \mathcal{N}^r/(n)} h_{nk} P_{nk}^r + W N_0}, \tag{1}$$

Where P_{nk}^r is defined as the transmitted power from access node n to UE k on resource block $r \in \mathcal{R}$, $\mathcal{R}$ is the total number of resource blocks (RBs), W is the bandwidth of each RB, N_0 is the thermal noise spectral power, $\sum_{n \in \mathcal{N}^r/(n)}$ are the access nodes which are using the resource block r and causing interference on the UE k associated with the access node n, and h_{nk} is the channel between access node n and the UE k. The channel model incorporates the effects of small-scale fading and large-scale fading (the latter includes path loss and shadowing).

The achievable downlink data rate at the UE k from AN n on single resource block r is given by

$$R_{nk}^r = W \log_2(1 + \gamma_{nk}^r). \tag{2}$$

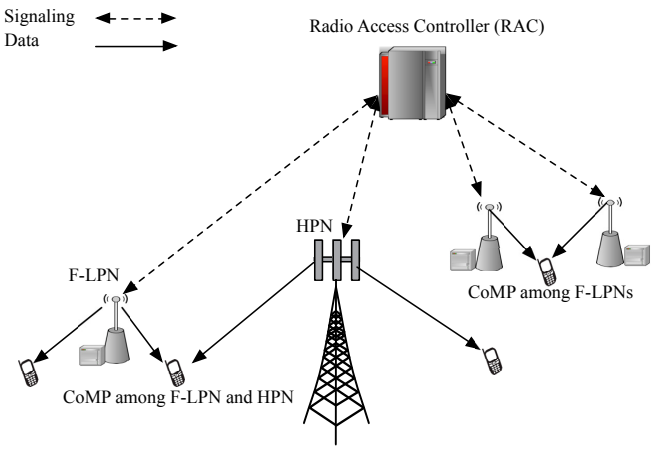

Fig. 1. System Model: Two-tier 5G HetNet consisting of RAC, HPN and F-LPNs.

Hence, the total achieved data rate from all the access nodes associated with UE k can be written as

$$\mathbb{R}_k = \sum_{n \in \mathcal{N}} \alpha_k^n \sum_{r \in \mathcal{R}} \psi_{nk}^r R_{nk}^r, \tag{3}$$

where $\alpha_k^n \in \{0, 1\}$ and $\psi_{nk}^r \in \{0, 1\}$ are the user association and the resource allocation control variables respectively.

To meet the data rate requirement, we assume that the UE k can be associated with M access nodes, which is referred to as CoMP transmission

$$\sum_{n \in \mathcal{N}} \alpha_k^n \leq M. \tag{4}$$

Thus, the AN n serves the UE k if the minimum data rate requirement in the service class tuple is guaranteed. Hence we can write

$$\mathbb{R}_k = \sum_{n \in \mathcal{N}} \alpha_k^n \sum_{r \in \mathcal{R}} \psi_{nk}^r R_{nk}^r \geq \eta_k. \tag{5}$$

The access nodes assign the needed resource blocks to UE k to satisfy (5). The number of resource blocks is given by

$$\Gamma = \left\lceil \frac{\eta_k}{\mathbb{R}_k} \right\rceil = \left\lceil \frac{\eta_k}{\sum_{n \in \mathcal{N}} \alpha_k^n \sum_{r \in \mathcal{R}} \psi_{nk}^r R_{nk}^r} \right\rceil, \tag{6}$$

where $\lceil \cdot \rceil$ is the ceiling function. Similarly, to meet the latency requirement of UE k

$$\frac{1}{\Delta_{nk}} \leq \tau_k, \tag{7}$$

where Δ_{nk} is the inverse of the statistical round trip delay-time (RTT) from access node n to UE k[1]. As mentioned previously, the F-LPNs are fitted with computation and caching capabilities that are brought close to the network edge. Therefore, the statistical RTT of the F-LPN is assumed to be much smaller than the statistical RTT of the HPN $\Delta_{nk}^{\text{F-LPN}} > \Delta_{nk}^{\text{HPN}}$ [15].

[1]The time required by a processor to serve a request is not included. It is assumed that with the advancement in computing powers and by aggregating multiple CPUs in a central unit, the processing latency can be neglected compared to the other latency components.

III. Problem Formulation And Proposed Bayesian Game Approach

A. Probelem Formulation

In this subsection, the problem of proper cell selection / user association is formulated. The utility functions are defined to represent the network's resource utilization and the UE's degree of satisfaction with the Quality of Service (QoS) including latency. The aim is to maximize the UE's utility function for improved QoS satisfaction and the network's resource utilization with respect to it's preferences. Therefore, the UE and the network utility functions, are respectively written as

$$U_k^{\text{UE}}(\alpha, \psi) = \sum_{n \in \mathcal{N}} \alpha_k^n \sum_{r \in \mathcal{R}} \psi_{nk}^r R_{nk}^r - \theta_k \Delta_{nk}, \tag{8}$$

$$U^{\text{Net}} = \omega_n(\theta_k) \left\lceil \frac{\Gamma_n}{\sum_{n \in \mathcal{N}} \alpha_k^n \mathbb{N}_n} \right\rceil, \tag{9}$$

where $\mathbb{N}_n$ is the total number of available resources in AN n and $\omega_n(\theta_k)$ is representing the AN n prefernce for UE k traffic type θ_k. The objective of the UE is to maximize (8) as

$$\begin{aligned} \max_{\alpha, \psi} : \quad & U_k^{\text{UE}}(\alpha, \psi), \\ \text{subject to} \quad & (4), (5), (6), (7). \end{aligned} \tag{10}$$

Similarly, the objective of the network is to maximize (9) as

$$\begin{aligned} \max_{\omega} : \quad & U^{\text{Net}}, \\ \text{subject to} \quad & \alpha_k^n, \psi_{nk}^r, \eta_k < \mathcal{C}, \end{aligned} \tag{11}$$

where $\mathcal{C}$ is the cap on the data rate allowed for UE k by access node n . We modeled the cell selection / user association problem as a Bayesian Game[2] [16]. The motivation for selecting a Bayesian game to model this problem is as follows. First, performing a joint optimization for both the UE and the network using regular optimization methods can be complex and computationally intensive. Also, as previously mentioned, some of the proposed solutions such as the densification in deploying LPNs and bringing the content closer to the user, are designed to accommodate the expected increase in data rates and to lower the end-to-end latency for certain use cases.

However, achieving the optimum system performance may not be feasible by just applying those solutions due to the lack of AN's *a priori* information about the UE's exact traffic type (at the access network, rather at the core network-level). Furthermore, assuming that the radio access nodes know the exact type of the UE's traffic in real-time via the core network level is unrealistic since this takes considerable time and makes achieving low latency quite unlikely. Hence, selecting a Bayesian game to model this problem, where perfect knowledge about UEs traffic types is not available at the access node, is well justified.

[2]The proposed cell selection / user association Bayesian Game algorithm is implemented in a central unit hypervisor referred to as the Radio Access Controller (RAC) which manages multiple access nodes.

B. Proposed Bayesian Game Approach

Definition: The cell selection / user association Bayesian game is defined in the strategic form as $\mathcal{G} = (\mathcal{P}, \Theta, \mathcal{A}, \mathbb{P}, U)$, where:

- **Players ($\mathcal{P}$):** Set of two players $\mathcal{P} = \{u_k, n_n\}$: the UE k and the AN n, respectively.
- **Types (Θ):** Set of possible types for UE k according to its traffic $\Theta = \{\theta_{k,1} = DS, \theta_{k,2} = DT\}$, where $\theta_{k,j}$ is the type j for the UE k.
- **Actions ($\mathcal{A}$):** The space of all possible combinational actions $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$. Where, the UE's action space $\mathcal{A}_1 = \{H = \text{UE selects HPN}, L = \text{UE selects LPN}\}$ and the AN's action space $\mathcal{A}_2 = \{S = \text{serve}, C = \text{CoMP}\}$.
- **Prior Probabilities ($\mathbb{P}$):** The probability $\mathbb{P}(\theta_j)$ over the types of users.
- **Utility functions (U):** The UE and the network utility functions as defined in (8) and (9), respectively.

The player i's strategy is mapping $s_i : \Theta_i \rightarrow \mathcal{A}_i$, which represents the player i's action for each possible type. In our Bayesian model, we started with an unbiased prior probability over the UE types. However, a given AN only knows its type and strategy, and does not know the strategies selected by the UEs or their actual traffic type. The expected utility of a player i under strategy profile $\mathbf{s}$ is

$$\mathbb{E}(U_i) = \sum_{\theta_j \in \Theta} U_i(s(\theta_j), \theta_j) \mathbb{P}(\theta_j). \tag{12}$$

Hence, the expected utility (payoff) of UE is

$$\begin{aligned} \mathbb{E}(U_k^{\text{UE}}) = \mathbb{P}(\theta_j) \sum_{n \in \mathcal{N}} \alpha_k^n \sum_{r \in \mathcal{R}} \psi_{nk}^r R_{nk}^r - \theta \Delta_{nk} \\ + (1 - \mathbb{P}(\theta_j)) \sum_{n \in \mathcal{N}} \alpha_k^n \sum_{r \in \mathcal{R}} \psi_{nk}^r R_{nk}^r - \theta \Delta_{nk}. \end{aligned} \tag{13}$$

Similarly, the expected utility (payoff) of network is

$$\begin{aligned} \mathbb{E}(U^{\text{Net}}) = \mathbb{P}(\theta_j) \left(\omega_n(\theta) \left\lceil \frac{\Gamma_n}{\sum_{n \in \mathcal{N}} \alpha_k^n \mathbb{N}_n} \right\rceil \right) \\ + (1 - \mathbb{P}(\theta_j)) \left(\omega_n(\theta) \left\lceil \frac{\Gamma_n}{\sum_{n \in \mathcal{N}} \alpha_k^n \mathbb{N}_n} \right\rceil \right). \end{aligned} \tag{14}$$

The UE k is of type $\theta_{k,1}$, when $U_k^{\text{UE}}(L, S) > U_k^{\text{UE}}(H, S)$ and $U_k^{\text{UE}}(L, C) > U_k^{\text{UE}}(H, C)$. Consequently strategy L strictly dominates strategy H. On the other hand, the UE is of type $\theta_{k,2}$, when $U_k^{\text{UE}}(H, S) > U_k^{\text{UE}}(L, S)$ and $U_k^{\text{UE}}(H, C) > U_k^{\text{UE}}(L, C)$. Then, strategy H strictly dominates strategy L. Starting with an unbiased belief (*a priori* probability) about the types of the UEs, then the expected utility (payoff) of the dominant strategies[3] for both traffic types is assumed to be even (at the beginning the network is not biased in its choice for any type of users). Hence, using TABLE I, the posterior probability P can be calculated as in (15).

TABLE I
PAYOFF MATRIX

Delay Sensitive (DS)			Delay Tolerant (DT)	
P			1-P	
UE Connects to	Network		Network	
	S	C	S	C
H	$(U_k^{\text{UE}}(H,S), U^{\text{Net}})$	$(U_k^{\text{UE}}(H,C), U^{\text{Net}})$	$(U_k^{\text{UE}}(H,S), U^{\text{Net}})$	$(U_k^{\text{UE}}(H,C), U^{\text{Net}})$
L	$(U_k^{\text{UE}}(L,S), U^{\text{Net}})$	$(U_k^{\text{UE}}(L,C), U^{\text{Net}})$	$(U_k^{\text{UE}}(L,S), U^{\text{Net}})$	$(U_k^{\text{UE}}(L,C), U^{\text{Net}})$

$$P = \frac{\left\lceil \frac{\Gamma_n}{\sum_{n\in\mathcal{N}} \alpha_k^n \mathbb{N}_n} \right\rceil - \omega_{n_0}(\theta_{k,j}) \left\lceil \frac{\Gamma_{n_0}}{\alpha_k^n \mathbb{N}_{n_0}} \right\rceil}{\omega_n(\theta_{k,j}) \left\lceil \frac{\Gamma_{\mathcal{N}\setminus n_0}}{\alpha_k^{\mathcal{N}\setminus n_0} \mathbb{N}_{n_0}} \right\rceil - \omega_{n_0}(\theta_{k,j}) \left\lceil \frac{\Gamma_{n_0}}{\alpha_k^n \mathbb{N}_{n_0}} \right\rceil - \left\lceil \frac{\Gamma_n}{\sum_{n\in\mathcal{N}\setminus n_0} \alpha_k^n \mathbb{N}_n} \right\rceil + \left\lceil \frac{\Gamma_n}{\sum_{n\in\mathcal{N}} \alpha_k^n \mathbb{N}_n} \right\rceil}. \quad (15)$$

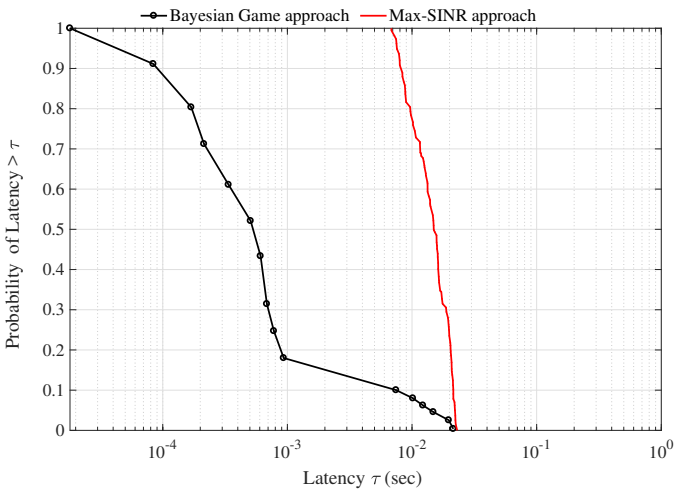

Fig. 2. CCDFs of achieved latency for delay-sensitive traffic

IV. SIMULATION RESULTS AND ANALYSIS

In this section, we present the simulation setup and discuss the performance results of the proposed Bayesian cell selection / user association algorithm. The performance of the proposed Bayesian cell selection approach is evaluated in terms of the probability of proper association and the achieved latency with respect to conventional CRE and Max-SINR based cell selection / user association algorithms used in LTE-Advanced. In the simulation, proper association is defined as the average number of delay-sensitive and delay-tolerant traffic associated with LPNS and HPN, respectively.

The simulation parameters are presented in TABLE II. In the simulation setup, we assume that there is one HPN and F-LPNs that are randomly deployed over the service region. The UEs are randomly distributed following a homogeneous *PPP* and each UE could be one of the two types DS or DT. The DT and DS applications traffic demands are modeled as a uniform random variable on [5,10] *Mbps* and [0.1,4] *Mbps*, respectively. The path loss is based on TS 36.942 model [17], for an urban environment as

$$PL = 40(1 - 4\times 10^{-3} hb)\log_{10}(d) - 18\log_{10}(hb) + 21\log_{10}(f) + 80 \text{ dB}, \quad (16)$$

where d is the separation between the UE and the AN, hb the height of the AN antenna and the f is the carrier frequency. The shadow fading is lognormal distributed and the fast fading is based on the Winner II model [18]. In simulation, the statistical RTT for the F-LPN $(1/\Delta_{nk}^{\text{F-LPN}})$ is modeled as a uniform random variable on [0.5,1.5] *ms* and for HPN $(1/\Delta_{nk}^{\text{HPN}})$ as the sum of three uniform random variables, namely: radio access [0.5,1.5] *ms*, the core network [1,2] *ms* and [5,10] *ms* for the Internet.

After performing simulations, we used the CCDF of achieved latencies of the delay-sensitive traffic to evaluate the Bayesian game in comparison with the Max-SINR approach. As shown in Fig. 2, the minimum achieved latency using the SINR approach is $\approx$ 6.8 *ms* due to the improper association that happens when delay-sensitive traffic is associated with HPN using the max-SINR criterion while the Bayesian approach achieves latencies $<$ 1 *ms*.

In addition to latency, we study the proper association of different schemes. Fig. 3 illustrates the cumulative distribution functions (CDFs) of proper associations for the proposed Bayesian game compared with the CRE and Max-SINR approaches. It is noted from Fig. 3 that the minimum percentage of proper associations for the delay-sensitive traffic of the Bayesian approach is greater than or equal to 82%, which outperforms the proper association of the conventional Max-SINR and the CRE approaches. Similarly, using the CDF of the delay-tolerant traffic as illustrated in Fig. 4, we observed that the CRE and Max-SINR approaches fall behind the Bayesian game. The Bayesian game approach attains a proper association greater than or equal to 80% for delay-tolerant traffic.

V. CONCLUSION

The three solutions proposed for 5G, namely: Dense LPNs, F-RANs and CoMP are not sufficient in themselves to meet the

[3]In game theory, player i's strategy is called a dominant strategy if it has a higher payoff than the payoff of all other strategies, that is $U_i^* > U_i^j$, regardless of the actions of the other players.

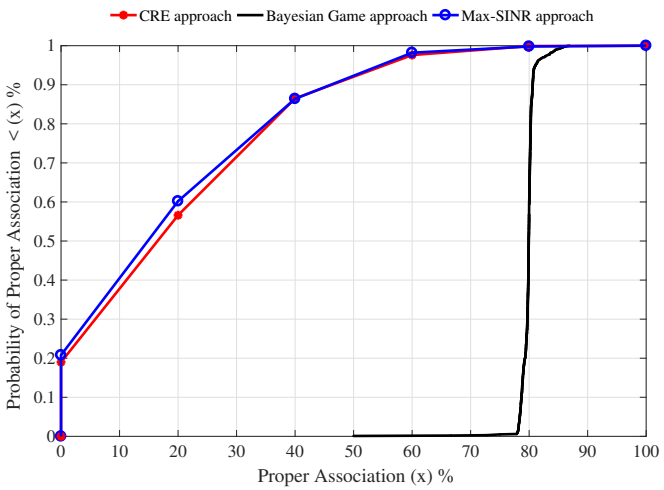

Fig. 3. CDFs of proper association for delay-sensitive traffic

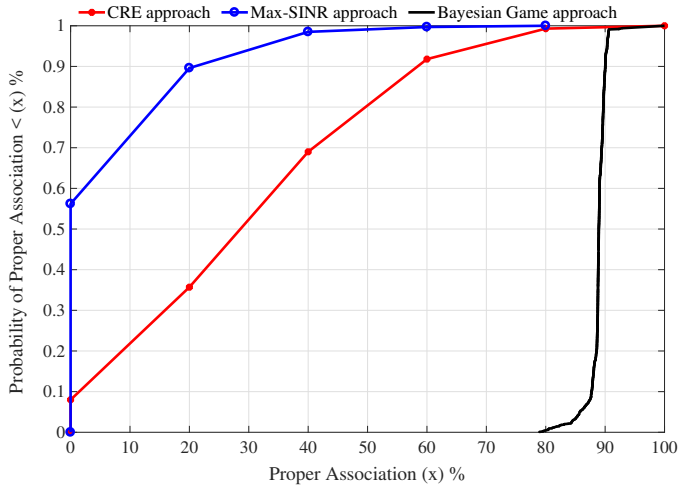

Fig. 4. CDFs of proper association for delay-tolerant traffic

diverse requirements for a wide range of emerging applications and in particular the low-latency target. For this reason, it is essential to rethink the traditional cell selection/ user association procedures, by considering both the traffic type and the access node capabilities. In this paper, a novel method for cell selection/ user association for 5G heterogeneous networks is proposed using Bayesian game. The utility functions of the user equipment and the network are defined based on the achievable data rate, the statistical RTT and the access node's traffic preferences. Simulation results demonstrate that the proposed Bayesian game algorithm provides a significant improvements in terms of the probability of proper association (as a function of traffic type) and the achieved latency compared with Max-SINR criterion and CRE approach. Such a methodology can be quite important in achieving the 5G low latency objective.

TABLE II
SIMULATION PARAMETERS

Parameter	Value
Service area	400 m × 400 m
Number of ANs $\mathcal{N} = n_0 \cup \mathcal{L}$	10 = 1+9
F-LPNs density (λ)	20 F-LPN/km^2
Number of UEs per AN	40
Total transmit power of ANs	{46.02,40} dBm
Transmit antenna height of ANs (*hb*)	20 m
Carrier frequency (*f*)	2.14 GHz
Bandwidth	20 MHz

ACKNOWLEDGMENT

The authors thank Dr. Eren Balevi for the fruitful discussion and valuable suggestions.

REFERENCES

[1] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: next-generation wireless broadband technology [invited paper]," *IEEE Wireless Communications*, vol. 17, no. 3, pp. 10–22, June 2010.

[2] 3GPP, "Coordinated Multi-Point Operation for LTE Physical Layer Aspects (Release 11)," TR 36.819, December 2011.

[3] J. G. Andrews *et al.*, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[4] 3GPP, "Service requirements for next generation new services and markets; stage1 (release 15)," TS 22.261, August 2016.

[5] A. Checko *et al.*, "Cloud RAN for mobile networks;a technology overview," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 405–426, Firstquarter 2015.

[6] C. L. I, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: a 5G perspective," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 66–73, February 2014.

[7] M. Peng, Y. Li, J. Jiang, J. Li, and C. Wang, "Heterogeneous cloud radio access networks: a new perspective for enhancing spectral and energy efficiencies," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 126–135, December 2014.

[8] A. Sengupta, R. Tandon, and O. Simeone, "Cache aided wireless networks: Tradeoffs between storage and latency," in *2016 Annual Conference on Information Science and Systems (CISS)*, March 2016, pp. 320–325.

[9] M. Peng, S. Yan, K. Zhang, and C. Wang, "Fog-computing-based radio access networks: issues and challenges," *IEEE Network*, vol. 30, no. 4, pp. 46–53, July 2016.

[10] Y. Y. Shih, W. H. Chung, A. C. Pang, T. C. Chiu, and H. Y. Wei, "Enabling low-latency applications in fog-radio access networks," *IEEE Network*, vol. 31, no. 1, pp. 52–58, January 2017.

[11] T. C. Chiu, W. H. Chung, A. C. Pang, Y. J. Yu, and P. H. Yen, "Ultra-low latency service provision in 5G fog-radio access networks," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2016, pp. 1–6.

[12] S. Ahmadi, *LTE-Advanced: A Practical Systems Approach to Understanding 3GPP LTE Releases 10 and 11 Radio Access Technologies*, 10 2013.

[13] 3GPP, "Range Expansion Techniques for HetNets," R1 124530, Qualcomm Incorporated, October 2012.

[14] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, June 2013.

[15] GSMA and Intelligence, "Understanding 5G: Perspectives on future technological advancements in mobile," *GSMA Intelligence Understanding 5G*, no. December, pp. 3–15, 2014.

[16] Y. Shoham and K. Leyton-Brown, *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge Univ. Press, 2008.

[17] 3GPP, "E-UTRA; LTE RF system scenarios," TS 36.942, 2008.

[18] P. Kysti, J. Meinil *et al.*, "IST-4-027756 WINNER II," EBITG,TUI, UOULU, CU/CRC, NOKIA, Tech. Rep., September 2007.