INVITED PAPER	*Special Section on Communication Technologies and Service Qualities in Various Access Networks*

User Pre-Scheduling and Beamforming with Imperfect CSI for Future Cloud/Fog-Radio Access Networks*

Megumi KANEKO[†a], *Member*, **Lila BOUKHATEM**[††b], **Nicolas PONTOIS**[†], *Nonmembers*, *and* **Thi-Hà-Ly DINH**[†c], *Student Member*

SUMMARY   By incorporating cloud computing capabilities to provide radio access functionalities, Cloud Radio Access Networks (CRANs) are considered to be a key enabling technology of future 5G and beyond communication systems. In CRANs, centralized radio resource allocation optimization is performed over a large number of small cells served by simple access points, the Remote Radio Heads (RRHs). However, the fronthaul links connecting each RRH to the cloud introduce delays and entail imperfect Channel State Information (CSI) knowledge at the cloud processors. In order to satisfy the stringent latency requirements envisioned for 5G applications, the concept of Fog Radio Access Networks (FogRANs) has recently emerged for providing cloud computing at the edge of the network. Although FogRAN may alleviate the latency and CSI quality issues of CRAN, its distributed nature degrades network interference mitigation and global system performance. Therefore, we investigate the design of tailored user pre-scheduling and beamforming for FogRANs. In particular, we propose a hybrid algorithm that exploits both the centralized feature of the cloud for globally-optimized pre-scheduling using imperfect global CSIs, and the distributed nature of FogRAN for accurate beamforming with high quality local CSIs. The centralized phase enables the interference patterns over the global network to be considered, while the distributed phase allows for latency reduction, in line with the requirements of FogRAN applications. Simulation results show that our proposed algorithm outperforms the baseline algorithm under imperfect CSIs, jointly in terms of throughput, energy efficiency, as well as delay.

key words: *5G and beyond, CRAN, FogRAN, user clustering, beamforming, radio resource management*

1. Introduction

Future dominating scenarios of human-centric communications, in conjunction with an ever increasing number of communicating devices, will generate a huge volume of mobile and wireless data traffic. This massive data is intended to be supported by the fifth generation (5G) communication system under severe spectrum deficiencies, while satisfying more stringent user Quality of Service (QoS) and Quality of Experience (QoE) levels. To meet these requirements, the 5G system will rely on several key enabling technologies among which Cloud Radio Access Networks (CRANs), that consider incorporating cloud computing capabilities to provide radio access functionalities [2].

In a CRAN, Remote Radio Heads (RRHs) performing only basic physical layer tasks such as Radio Frequency (RF) and A/D conversion are geographically deployed to cover several small cells. The users' RF signals collected by the RRHs are transmitted to the cloud platform through fronthaul links. A centralized server that groups the cloud Baseband Units (BBUs) enables to perform the radio access tasks and signal processing. Such a CRAN system is able to adapt to non-uniform traffic and to utilize the resources more efficiently. Due to the simplicity of RRHs and their easy deployment compared to their counterparts in legacy cellular systems, CRAN has also the potential to decrease the cost of network operation by reducing power and energy consumption. In addition, CRAN enables optimal joint baseband signal processing, radio resource allocation and interference management but at the expense of heavily burdening the capacity-limited fronthaul links [3], [4]. Hence, many CRAN related research works have focused on the optimization of user clustering and beamforming under fronthaul capacity constraints [5]–[10]. Another major drawback of the CRAN centralized architecture is the additional network latency introduced by the fronthaul links, making it unsuited for the highly delay-sensitive applications envisioned in 5G. The induced delay also entails outdated and hence imperfect Channel State Information (CSI) knowledge at the cloud side of the link qualities between all APs and users, causing important performance degradation of resource allocation and beamforming schemes in CRAN [11].

Recently, the Edge Computing (EC) concept has emerged as another 5G architectural feature that pushes mobile computing, system control and storage toward network edges. These edge nodes are expected to provide sufficient computational resources, energy and latency reductions for satisfying the latency-critical applications as well as the offloading requirements of resource-limited end-devices. Several Edge Computing solutions have been considered by both research and industry such as Mobile Edge Computing (MEC), Cloudlets or Fog Radio Access Networks (FogRAN) [12]. In these solutions, some of the physical and MAC layer BBU functionalities are shifted from cloud BBUs

to edge nodes, as well as part of the computing and caching capabilities. This structure is expected to drastically alleviate the burden on fronthaul links and to meet the stringent delay requirements of edge users [13], but at the cost of lower network-wide optimality. Several works have exploited the edge processing capability to enhance the performance of various applications, and analyzed the joint optimization of cloud and edge processing [14]. However, very few works have addressed the design of optimized physical and MAC layers under the novel FogRAN architecture, that incorporate the heterogeneous CSI qualities at the cloud and edge APs. This is a key problem since optimized lower layers, in particular user clustering and beamforming, will have a huge impact on the actual performance of FogRANs at the application level.

In this paper, we investigate the joint user clustering and beamforming issues in FogRAN, which are fundamental enabling elements for future edge computing applications. Our proposed solution has the advantage of considering FogRAN specific network and information constraints: i) the distributed and limited computational capabilities of FogAPs with only local channel knowledge, and ii) the fronthaul constraints between the cloud and each FogAP. More specifically, we consider the downlink weighted sum-rate maximization problem. Our proposed mechanism includes two complementary centralized and distributed actions that use both imperfect and perfect CSI knowledge. First, the proposed scheme carries out a centralized user pre-scheduling that provides the optimal user clustering to each FogAP, taking into account all interferences based on global but imperfect CSI knowledge, due to the transport delays compelled by fronthaul links. Then, beamforming vectors are computed at each FogAP for its own pre-scheduled users, using perfect CSI knowledge since the delay of CSI feedback is negligible compared to the fronthaul delays. We also define different levels of CSI knowledge at each FogAP given different CSI feedback strategies from users to each FogAP, namely local, intermediate, and global perfect CSI knowledge. The impact of each level of CSI knowledge at FogAPs on the global network performance is analyzed and discussed. The main contributions of our work are listed as follows:

1) The proposed scheme jointly exploits the centralized cloud processing for large-scale user clustering and distributed local beamforming, given the heterogeneous CSI qualities imposed by FogRAN. The scheme accounts for the fact that beamforming is highly sensitive to CSI accuracy, unlike user clustering.

2) For cloud centralized pre-scheduling, we formulate a weighted sum-rate maximization under the FogRAN-specific constraint where only one FogAP is allowed to serve each user in a scheduling period, as pointed out in [12]. This gives rise to a difficult non-convex optimization problem with a discrete constraint, for which we propose a relaxation method.

3) To enable low-complexity and distributive sum-rate maximization at each FogAP, Signal to Leakage-plus-Noise Ratio (SLNR) maximization is chosen as it approaches Signal to Interference-plus-Noise Ratio (SINR) maximization under local CSI [15].

4) The numerical evaluations show that, compared to centralized CRAN, the proposed method achieves much lower latencies for higher sum-rate and energy efficiency, in the presence of imperfect CSIs. The proposed design hence ensures an essential trade-off between the network-wide utility performance, required computational complexity at FogAPs and packet latencies.

2. System Model

2.1 CRAN and FogRAN Architectures for Core or Edge Intelligence

We consider two types of architectures referred as CRAN and FogRAN depending on the intelligence location either towards the core or edge of the network, as depicted in Fig. 1. In the CRAN case, we assume a centralized system where all the signal processing and resource management tasks are performed at the cloud BBU pool. R Macro or Pico RRHs (APs) in set $\mathcal{R}$ are connected to the cloud through fronthaul links of respective capacities C_r. Each AP r is equipped with M_r transmit antennas. The set of all mobile users is denoted by $\mathcal{K}$ with cardinality K. Each user terminal is equipped with one receive antenna. We denote by $\mathbf{w}_{rk} \in \mathbb{C}^{M_r \times 1}$ the beamforming vector of AP r to user k. The concatenated beamforming vector of all AP antennas is defined as $\mathbf{w}_k = [\mathbf{w}_{1k}^H, \mathbf{w}_{2k}^H, \cdots, \mathbf{w}_{Rk}^H]^H \in \mathbb{C}^{M \times 1}$ for user k, where $M = \sum_{r \in \mathcal{R}} M_r$ is the total number of transmit antennas and $(.)^H$ denotes Hermitian transpose. Similarly, $\mathbf{h}_{rk} \in \mathbb{C}^{M_r \times 1}$ is the channel vector between AP r and user k and $\mathbf{h}_k = [\mathbf{h}_{1k}^H, \mathbf{h}_{2k}^H, \cdots, \mathbf{h}_{Rk}^H]^H \in \mathbb{C}^{M \times 1}$ the channel vector from all APs to user k. The received signal y_k by user k is given by

$$y_k = \mathbf{h}_k^H \mathbf{w}_k s_k + \mathbf{h}_k^H \sum_{\substack{k' \in \mathcal{K} \\ k' \neq k}} \mathbf{w}_{k'} s_{k'} + n_k, \tag{1}$$

where s_k is the transmit message for user k drawn independently from the signal constellation with zero mean and unit
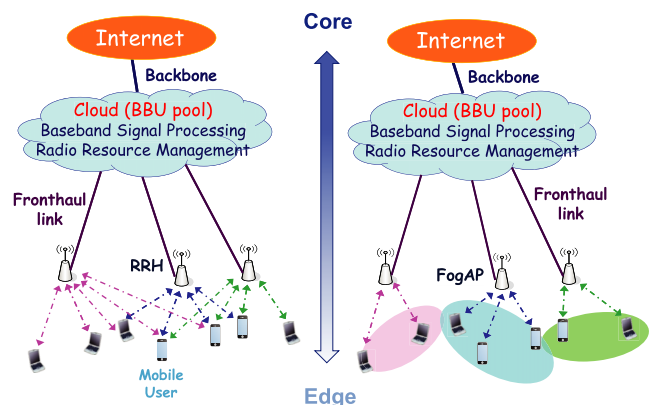

Fig. 1 CRAN (left) and FogRAN (right) architectures.

variance, and $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ denotes the AWGN noise where σ_n^2 is the noise power. The first term in (1) is the desired signal, and the second is the interference resulting from the other users' signals. The beamforming vectors $\mathbf{w}_k$ will be optimized at the cloud BBUs for all users, and any user may be served by any of the R APs.

In the FogRAN architecture, the intelligence is pushed towards the edge by enhancing traditional RRHs with higher processing capabilities, allowing basic signal processing tasks. Therefore, for sake of differentiation these RRHs will be referred as FogAPs (Macro or Pico) as in Fig. 1. In our proposed scheme, the beamforming vectors will be optimized locally at each FogAP r. The received signal of user k served by FogAP r is also given by (1), but where in $\mathbf{w}_k = [\mathbf{w}_{1k}^H, \mathbf{w}_{2k}^H, \cdots, \mathbf{w}_{Rk}^H]^H$, only the beamforming vector $\mathbf{w}_{rk}$ that corresponds to the Fog AP r associated to user k is non-zero, in line with the FogRAN requirement described in [12]. The set of users associated to FogAP r is denoted $\mathcal{K}_r$ with cardinality K_r.

From (1), the SINR of user k is written as

$$\gamma_k = \frac{\left| \mathbf{h}_k^H \mathbf{w}_k \right|^2}{\sum_{\substack{k' \in \mathcal{K} \\ k' \neq k}} |\mathbf{h}_k^H \mathbf{w}_{k'}|^2 + \sigma_n^2}. \tag{2}$$

The achievable rate for user k is thus given by

$$R_k = B \log_2(1 + \gamma_k), \tag{3}$$

where B denotes the communication bandwidth.

2.2 Imperfect but Global CSI at Cloud BBUs

In centralized CRAN, optimal resource allocation can be performed using global CSI knowledge, i.e., all channel vectors $\mathbf{h}_{rk}$ for all APs r and all users k. However, the fronthaul links will introduce non-negligible delays as pointed out in [12] causing imperfect global CSI. The stochastic error model will be assumed as in [11], [16], where the imperfect channel vector of global CSI is given by

$$\widetilde{\mathbf{h}}_{rk} = \mathbf{h}_{rk} + \mathbf{e}_{rk}, \tag{4}$$

where $\mathbf{e}_{rk} \sim \mathcal{CN}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{M_r})$ with σ_e^2 the component-wise power of the global CSI error. Then, the concatenated imperfect CSI is defined as $\widetilde{\mathbf{h}}_k = [\widetilde{\mathbf{h}}_{1k}^H, \widetilde{\mathbf{h}}_{2k}^H, \cdots, \widetilde{\mathbf{h}}_{Rk}^H]^H \in \mathbb{C}^{M \times 1}$. Thus, only these outdated channels $\widetilde{\mathbf{h}}_k$ for all users k will be available at the BBUs, i.e., *global but imperfect CSI*. The same assumption will hold for the proposed allocation algorithm for FogRANs, during the pre-scheduling phase performed at the cloud BBUs.

2.3 Perfect but Partial CSI at Each FogAP

By contrast, in the FogRAN case, perfect knowledge of CSI $\mathbf{h}_{rk}$ will be assumed at each FogAP r, since the delay due to CSI feedback on the wireless links between each user and AP is negligible compared to the transport delays induced by the fronthaul links. We assume different levels of CSI knowledge at each FogAP, corresponding to different CSI feedback strategies from users to FogAPs:

- *Local perfect CSI:* each FogAP only knows the CSI levels of its own clustered users. It has no knowledge about the interference channels towards users served by different FogAPs. Therefore, this case incurs the lowest amount of CSI feedback overhead.
- *Global perfect CSI:* each FogAP knows all CSI levels of all users in the system. This unrealistic case that produces the maximum amount of CSI feedback overhead, is evaluated for comparison purposes.
- *Intermediate perfect CSI:* between the above two cases, it is assumed that each Macro FogAP knows the interference channel states towards users served by its neighboring FogAPs, in addition to its own associated users, at the cost of additional CSI feedback overhead compared to the *Local perfect CSI* case. In particular, each Macro FogAP has knowledge of CSIs of all its associated users and of users associated to Pico FogAPs located within its coverage, while Pico FogAPs have only knowledge of CSIs of their own associated users.

3. Reference Centralized Algorithm for CRAN

We focus on weighted sum-rate maximization subject to fronthaul constraints as in [5]. Optimal user clustering and beamforming vectors are determined at the BBU pool using global CSI. The optimization problem is formulated as

$$\max_{\mathbf{w}_{rk}, R_k} \sum_{k \in \mathcal{K}} \alpha_k R_k \tag{5a}$$

$$s.t. \quad \sum_{k \in \mathcal{K}_r} ||\mathbf{w}_{rk}||_2^2 \leq P_r, \quad \forall r \in \mathcal{R}, \tag{5b}$$

$$\sum_{k \in \mathcal{K}_r} R_k \leq C_r, \quad \forall r \in \mathcal{R}, \tag{5c}$$

$$R_k \leq B \log_2(1 + \gamma_k), \quad \forall k \in \mathcal{K}, \tag{5d}$$

where α_k are weight parameters to achieve different fairness levels among users. The first constraint is given by the maximum power for each AP r, the second one is the per-AP fronthaul rate constraint, and the third one expresses the achievable rate for each user k.

This is a non-convex optimization problem for which a weighted MMSE-based algorithm was proposed [5], [6]. The case with perfect CSI represents the ideal scenario in terms of system performance, but requires full CSI feedback for all users from each AP, resulting into a significant burden over bandwidth-limited fronthaul links. In reality, the CSI used for this optimization will be necessarily outdated due to the delays introduced by fronthaul links. Therefore, this reference algorithm for CRAN will be evaluated under different levels of global CSI imperfectness at the cloud.

4. Proposed Cloud-Aided Distributed Algorithm for FogRAN

In the proposed scheme, we split the joint resource allocation tasks: the user pre-scheduling carried out centrally at the cloud BBUs, and the beamforming optimization carried out locally at each FogAP. The pre-scheduling consists in a user clustering, where the BBU pool decides to which FogAP each user should be assigned at given time frames. This pre-scheduling is performed periodically, every T frames, based on imperfect gobal CSI due to fronthaul delays. Given the resulting user clustering, each FogAP performs beamforming in each frame, using perfect local CSI (a common and reasonable assumption under low mobility users), since delays due to local CSI feedback over wireless links are negligible compared to fronthaul delays. Since FogAPs are uncoordinated during this beamforming phase, the pre-scheduling needs to determine optimal user clusterings forming a partition $(\mathcal{K}_r)_{r \in \mathcal{R}}$ of the set of all users. This is in contrast with the CRAN user clustering in Sect. 3, where each user may be served by any AP. Note that some subsets $\mathcal{K}_r$ may be empty, i.e., some FogAPs may not have any scheduled user for given frames.

In addition, to fully exploit the FogRAN architecture, the global weighted sum-rate utility is optimized at the cloud BBUs, since it requires a significant amount of computational complexity. However, performing again sum-rate optimization for beamforming would entail too much burden on each FogAP. Instead, at each FogAP we propose to maximize the SLNR metric, known to approach the performance of SINR maximization, hence of sum-rate maximization, but with much lower computational complexity as shown in [15][17]. The details of each phase are given below.

4.1 Pre-Scheduling

For user pre-scheduling, we propose a modified version of the weighted sum-rate optimization in CRAN (5a) based on the imperfect global CSIs in Sect. 2, formulated as

$$\max_{\mathbf{w}_{rk}, R_k} \sum_{k \in \mathcal{K}} \alpha_k R_k \tag{6a}$$

$$s.t. \sum_{k \in \mathcal{K}_r} \|\mathbf{w}_{rk}\|_2^2 \le P_r, \quad \forall r \in \mathcal{R}, \tag{6b}$$

$$\sum_{k \in \mathcal{K}_r} R_k \le C_r, \quad \forall r \in \mathcal{R}, \tag{6c}$$

$$R_k \le B \log_2(1 + \gamma_k), \quad \forall k \in \mathcal{K}, \tag{6d}$$

$$\sum_{r \in \mathcal{R}} \left\| \|\mathbf{w}_{rk}\|_2^2 \right\|_0 \le 1, \quad \forall k \in \mathcal{K}, \tag{6e}$$

where in the last constraint, the zero-norm follows $\|x\|_0 = 1$ if $x \ne 0$ and 0 otherwise. Thus, (6e) enforces that each user is associated to at most one FogAP, i.e., it ensures the partitioning required in FogRANs as mentioned above. This is a non-convex optimization problem which is difficult

to handle given the discrete constraint (6e). To solve this problem, we propose to transform constraint (6e) as follows, using a relaxation technique shown to be useful in different settings in [6], [18]. Namely, we introduce a parameter τ and define

$$\beta_{rk} = \frac{1}{\|\hat{\mathbf{w}}_{rk}\|_2^2 + \tau}, \tag{7}$$

where $\hat{\mathbf{w}}_{rk}$ denotes the value of $\mathbf{w}_{rk}$ at the algorithm's previous iteration. Constraint (6e) is then approximated by

$$\sum_{r \in \mathcal{R}} \left\| \|\mathbf{w}_{rk}\|_2^2 \right\|_0 \approx \sum_{r \in \mathcal{R}} \beta_{rk} \|\mathbf{w}_{rk}\|_2^2 \le 1. \tag{8}$$

The parameter τ is tuned such that $\beta_{rk} \|\mathbf{w}_{rk}\|_2^2$ approaches $\left\| \|\mathbf{w}_{rk}\|_2^2 \right\|_0$, throughout the iterations. Hence, we update at each iteration i,

$$\tau_i = \tau_0 \lambda^i, \tag{9}$$

with τ_0 an appropriate initial value of τ and $\lambda \in (0, 1)$.

After this transformation, problem (6a) is expressed in a form similar to the reference problem (5a), with an additional quadratic constraint. Hence, similarly to [6], [18], we can transform constraint (6d) into a convex form and derive a solution. The steps of the proposed algorithm are given in Algorithm 1.

Algorithm 1 Proposed Pre-Scheduling at BBUs

Initialize
   $\{\hat{\mathbf{w}}_{rk}, \tau\}$
repeat
1) Update β_{rk}
2) Update $\tau = \tau \cdot \lambda$
3) For fixed β_{rk}, cast (6a) as an SOCP and solve using
standard convex optimization tools.
until $\|\hat{\mathbf{w}}_{rk} - \mathbf{w}_{rk}\|_2^2 < \epsilon$

The obtained solutions give an implicit scheduling, so we can retrieve the user clustering as follows: $k \in \mathcal{K}_r$ if $\mathbf{w}_{rk} \ne \mathbf{0}$.

4.2 Local Beamforming

In order to efficiently optimize the local beamforming, we propose to maximize the SLNR of each user at each FogAP. SLNR optimization is especially suited in this case since FogAPs are unable to coordinate among themselves and do not have access to the global SINR levels experienced by their associated users. In addition, this optimization requires very low complexity which is vital for FogAPs, unlike the weighted sum-rate optimizations in (5a) or (6a) which require the high processing capabilities of cloud BBUs.

Given the level of perfect CSI knowledge at each FogAP described in Sect. 2.3, we define below the corresponding

levels of SLNR for user k:

- *Local perfect CSI:* denoted ζ_k^{Loc}, it is given as

$$\zeta_k^{\text{Loc}} = \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum\limits_{\substack{k' \in \mathcal{K}_r \\ k' \neq k}} |\mathbf{h}_{k'}^H \mathbf{w}_k|^2 + \sigma_n^2}, \tag{10}$$

where only the interference towards the FogAP's associated users $k' \in \mathcal{K}_r$ is taken into account,

- *Global perfect CSI:* denoted ζ_k^{Glo}, it is given as

$$\zeta_k^{\text{Glo}} = \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum\limits_{\substack{k' \in \mathcal{K} \\ k' \neq k}} |\mathbf{h}_{k'}^H \mathbf{w}_k|^2 + \sigma_n^2}, \tag{11}$$

where interference towards all users in the system $k' \in \mathcal{K}$ is considered,

- *Intermediate perfect CSI:* denoted ζ_k^{Int}, it is given as

$$\zeta_k^{\text{Int}} = \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum\limits_{\substack{k' \in \mathcal{K}_r^{\text{Neigh}} \\ k' \neq k}} |\mathbf{h}_{k'}^H \mathbf{w}_k|^2 + \sigma_n^2}, \tag{12}$$

where $\mathcal{K}_r^{\text{Neigh}}$ denotes the set of users associated to FogAP r and to its neighboring FogAPs. Hence, if r is a Macro FogAP, $\mathcal{K}_r^{\text{Neigh}} = \cup_{r' \in \text{Cov}(r)} \mathcal{K}_{r'}$ where $\text{Cov}(r)$ denotes the set of FogAPs r' included in the coverage of Macro FogAP r (including r itself), otherwise if r is a Pico FogAP: $\mathcal{K}_r^{\text{Neigh}} = \mathcal{K}_r$.

Thus, each FogAP r solves the following optimization problem for each associated user $k \in \mathcal{K}_r$. Here we assume equal power allocation of the FogAP power among its associated users. Depending on the CSI knowledge level expressed by variable $\phi \in \{\text{Loc}, \text{Glo}, \text{Int}\}$, the optimization problem at each FogAP is thus

$$\max_{\mathbf{w}_{rk}} \zeta_{rk}^{\phi} \quad \text{s.t.} \quad ||\mathbf{w}_{rk}||_2^2 \leq \frac{P_r}{K_r}. \tag{13}$$

The optimal beamforming vector solution can be obtained in closed-form solution according to [17]. For each CSI knowledge level, it is expressed as

$$\mathbf{w}_{rk}^{opt,\phi} = \sqrt{\frac{P_r}{K_r}} \max \text{eig} \left\{ \left(\sum_{k' \in \mathcal{K}_\phi,\, k' \neq k} \mathbf{h}_{rk'} \mathbf{h}_{rk'}^H + \frac{K_r \sigma_n^2}{P_r} \mathbf{I} \right)^{-1} \mathbf{h}_{rk} \mathbf{h}_{rk}^H \right\}, \tag{14}$$

where $\max \text{eig}(A)$ gives the eigenvector corresponding to the largest eigenvalue of matrix A. In (14), the summation set $\mathcal{K}_\phi$ is equal to $\mathcal{K}_r$, $\mathcal{K}$ and $\mathcal{K}_r^{\text{Neigh}}$ in the local, global, and intermediate CSI knowledge cases, respectively.

5. Numerical Results

5.1 Simulation Settings

We consider a wrap-around two-tier CRAN and FogRAN

to evaluate the reference and proposed algorithms. There are 3 Macro-RRHs (FogAPs) and 9 Pico-RRHs (FogAPs), where the transmit power of Macro and Pico-RRHs are 43 and 30 dBm respectively, and their fronthaul capacities C_r are set to (690,107) Mbps as in [6]. All channels are subject to Rayleigh fading and log-normal shadowing. The noise power spectral density is equal to -169dBm/Hz, and the bandwidth $B = 10$ MHz. Other system parameters also follow that of [6]. We evaluate the performance of the reference centralized weighted sum-rate optimization for CRAN in Sect. 3, denoted *CRAN (ref.)*, and the proposed pre-scheduling and local beamforming algorithm for FogRAN in Sect. 4, denoted *FogRAN (prop.)*, for $K = 60$ or 90 users and $\alpha_k = 1, \forall k$. For the proposed method, the pre-scheduling period was fixed to $T = 10$, unless stated otherwise. In all simulations, a full buffer traffic model was assumed, with one traffic flow per user. The scheduling frame duration is fixed to 1 ms. All results have been averaged over 3000 random user positions and channel realizations, each realization corresponding to a frame transmission.

5.2 Sum-Rate and User-Rate Performance

Figure 2 shows the network sum-rate degradation against different levels of the global CSI qualities available at the cloud BBUs, given by the CSI error variance σ_e^2 defined in Sect. 2.3. This degradation is compared to the performance under perfect global CSI, given by dotted lines for each algorithm. Note that both *CRAN (ref.)* and *FogRAN (prop.)* use the same imperfect global CSI at the cloud, while *FogRAN (prop.)* reuses the perfect local CSI available at the local FogAP, for beamforming only. Clearly, the centralized algorithm offers very high throughput for near-perfect global CSI, but degrades rapidly as the error variance grows. By contrast, the proposed algorithm for FogRAN shows a throughput loss due to the distributed beamforming for high quality global CSI, but also a much higher robustness against CSI errors and even a close to optimal performance for $\sigma_e^2 = 1$. Note that CSI errors in the order of $\sigma_e^2 = 0.1$ correspond to realistic CSI qualities in usual cellular systems [16], [19]. Given
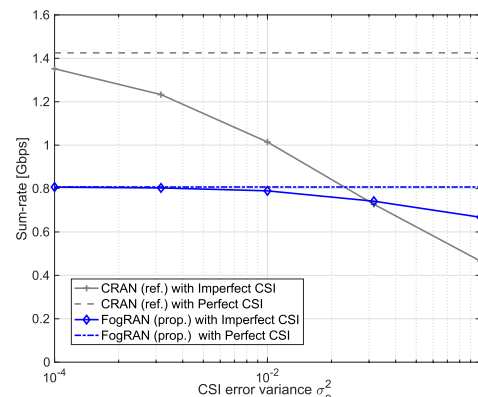

Fig. 2 Sum-rate performance of reference CRAN and proposed Fog RAN algorithms under global CSI imperfectness, $T = 10$, $K = 60$.
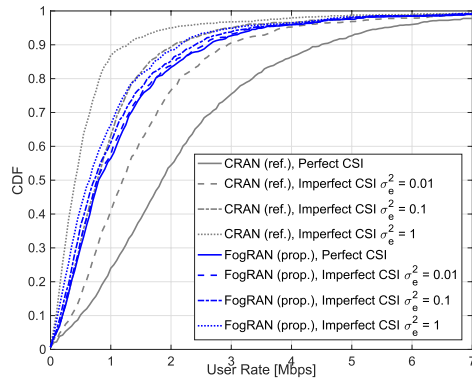
Fig. 3 CDF of per-user rates of reference CRAN and proposed Fog RAN algorithms under global CSI imperfectness, $T = 10$, $K = 60$.

Fig. 4 CDF of delays for reference and proposed algorithms under global CSI imperfectness, $P = 12$ kbits, $T = 10$, $K = 60$.

Fig. 5 CDF of delays for reference and proposed algorithms under global CSI imperfectness, $P = 1$ kbit, $T = 10$, $K = 60$.

the additional fronthaul delays, $\sigma_e^2 \geq 0.1$ hence constitutes a realistic region of interest for the considered CRAN framework. Thus, for the CSI qualities of interest, the proposed algorithm is shown to outperform the reference one.

Figure 3 shows the Cumulative Distribution Function (CDF) of the user rates. It can be observed that, while centralized CRAN outperforms the proposed algorithm for perfect and near-perfect global CSI levels ($\sigma_e^2 = 0.01$), this tendency is reversed as σ_e^2 grows. In particular, for $\sigma_e^2 = 1$, the proposed scheme allows 90% of users to achieve up to 2.2 Mbps against 1.3 Mbps for the reference CRAN, hence a 70% increase. Moreover, the curves for different global CSI qualities show a much narrower spread compared to those of centralized CRAN. This results into much smaller variations of achievable user rates under CSI uncertainties, confirming the robustness of our scheme unlike the reference one.

5.3 Delay Performance

We evaluate the delay performance of these algorithms, one of the key aspects of FogRANs and edge computation. The delay here is defined as the time required for receiving a message length of P bits, measured over 3000 realizations and 60 users, where each realization corresponds to a scheduling frame of length 1 ms. No retransmissions are considered. Figures 4 and 5 show the CDF of delays, on one hand for a relatively large message length of $P = 12$ kbit, and on the other hand, a smaller message length of $P = 1$ kbit as in, e.g., control messages for IoT applications. Each discrete incremental step corresponds to an additional scheduling frame. First, we can see from Fig. 4 that for larger messages, the proposed scheme is outperformed by reference CRAN for perfect and near-perfect ($\sigma_e^2 = 0.01$) global CSI qualities. However, the advantage of the proposed scheme grows with σ_e^2, cutting down to half the delay at the 50th percentile for $\sigma_e^2 = 1$ (20 ms against 40 ms for the reference). Note that the wide spread of the curves for the reference scheme demonstrates the unsuitability of such fully centralized solution for supporting delay-stringent 5G applications. By contrast, the narrow spread of the curves for the proposed scheme also shows its robustness against CSI uncertainties in terms of
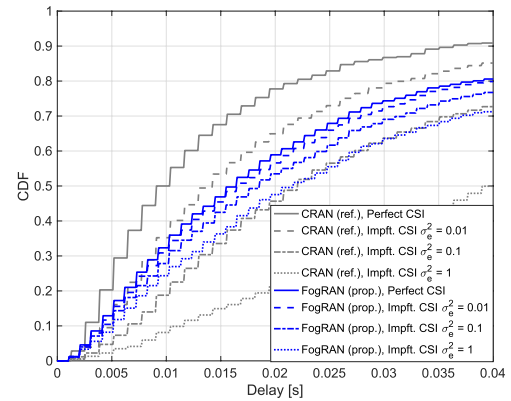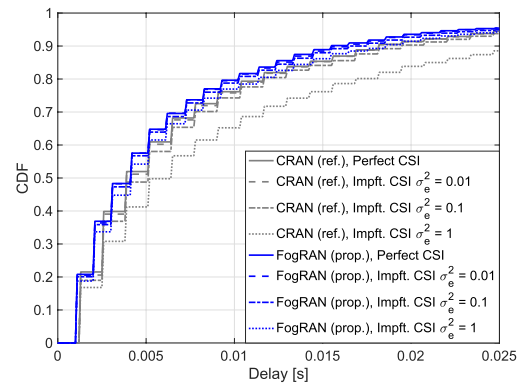
delay.

In the case of short messages of 1 kbit, Fig. 5 shows that the proposed algorithm outperforms the conventional one for all global CSI qualities, even in the perfect CSI case. This can be understood as follows. Even though the proposed scheme achieves lower total sum-rate, it enables to serve, through the distributed but accurate beamforming, rates that are high enough in order to receive small messages efficiently. On the contrary, the centralized scheme allows to boost the throughput by globally concentrating the resources towards the users with best channel conditions, at the detriment of users in lower conditions whose perceived delay is increased. Furthermore, as the CSI errors increase, the throughput of best users diminishes drastically, thereby degrading the delay performance of the centralized scheme.

5.4 Energy Efficiency

Next, we compare the different algorithms in terms of energy efficiency, defined as the ratio between the system sum-rate over the bandwidth B and the total consumed power,

$$\eta = \frac{B \sum_k R_k}{P_{\text{prop}}}. \tag{15}$$

Since the BBU processing occurs in 1 every T frames

Table 1 Parameters and values in the power consumption model.

Parameter	Value [20]	Description
$P_{\text{fix},r}$	0.825W	Traffic-independent fixed power consumption
ρ	4/3	Redundancy in the fronthaul transport interface
b_{IQ}	20	Number of IQ samples bits
f_{pre}	1.5MHz	Frequency of updating the precoder
$P_{\text{td},r}$	0.25W/Gbps	Traffic-dependent power
ξ_r	0.4	Efficiency of power amplifier
$P_{\text{ic},r}$	0.2W	Power for FogAP antenna circuit components

while only FogAP processing occurs in all other frames, the consumed power P_{prop} is given by

$$P_{\text{prop}} = \frac{1}{T}\Big[\Big(\sum_{r\in\mathcal{R}}(P_r^{f,u}+P_r^c+P_r^{f,d}+P_r^w)\Big)+\sum_{t=1}^{T-1}\sum_{r\in\mathcal{R}}(P_r^c+P_r^w)\Big],$$
(16)

where the first term expresses the power consumed during pre-scheduling in the first frame of the period, while the second term gives the power consumed at each FogAP during subsequent frames. From the models in [20], [21], we have

- $P_r^{f,u}$: power consumed for transmitting user CSIs from FogAP r to the cloud through the fronthaul link,
- P_r^c: power consumed by the power amplifier and circuit at FogAP r,
- $P_r^{f,d}$: power consumed on the fronthaul link of FogAP r by transmissions from the cloud,
- P_r^w: power consumed on the wireless links for downlink transmissions to its associated users in $\mathcal{K}_r$.

Given the parameters in Table 1, they can be determined as

$$P_r^{f,u} = P_{\text{fix},r} + \rho M_r K_r b_{IQ} f_{\text{pre}} P_{\text{td},r},$$
(17)

$$P_r^c = \frac{1}{\xi_r}BP_r + M_r P_{\text{ic},r},$$
(18)

$$P_r^{f,d} = P_{\text{fix},r} + B\sum_k R_k P_{\text{td},r},$$
(19)

$$P_r^w = \sum_{k\in\mathcal{K}_r} ||\mathbf{w}_{rk}||_2^2.$$
(20)

By contrast, the consumed power for the conventional CRAN algorithm is given by

$$P_{\text{conv}} = \sum_{r\in\mathcal{R}} P_r^{f,u}+P_r^c+P_r^{f,d}+P_r^w.$$
(21)

Figure 6 shows that the proposed FogRAN algorithm outperforms conventional CRAN in the case of imperfect CSI, with an increasing gain with the amount of CSI error. Even under perfect CSI, the proposed algorithm achieves a similar energy efficiency as compared to CRAN algorithm, despite a much lower sum-rate performance as seen in Fig. 2. This is explained by the significant savings of power consumption over the fronthaul links, thanks to the hybrid cloud
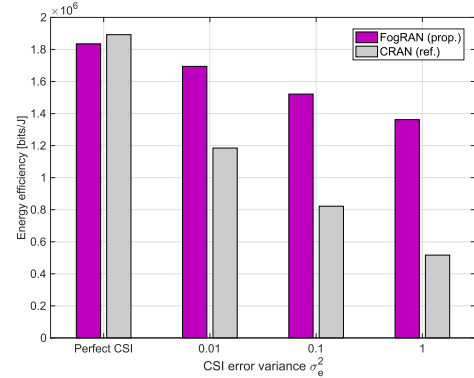
Fig. 6 Energy efficiency of reference and proposed algorithms, different levels of global CSI imperfectness, $T = 10$, $K = 60$.
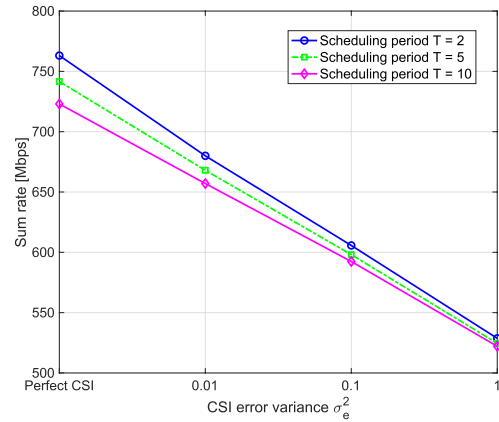
Fig. 7 Average sum-rate comparison of proposed algorithms for different values of the scheduling period T, for $K = 60$.

and FogAP processing structure of the proposed scheme.

5.5 Performance Against Varying Pre-Scheduling Period T

Comparing the proposed FogRAN algorithm for different values of the scheduling period T, Fig. 7 shows that as expected, smaller periods provide a higher sum-rate performance, given the higher frequency of clustering optimization. However, we can see that the performance gaps decrease for higher CSI errors. For $\sigma_e^2 = 1$, a period $T = 10$ is sufficient as it provides a very close performance to $T = 2$, since the imperfect CSIs are useless for improving the performance through optimized clustering. Hence, compared to conventional CRAN, at high CSI errors, a higher sum-rate may be achieved even with large periods, providing significant energy and computational savings.

5.6 Effect of CSI Knowledge at Each FogAP

Finally, we evaluate the effects of the different perfect CSI knowledges at each FogAP, described in Sect. 2.3. Figure 8 shows the total average sum-rate performance against CSI error variances. We observe that for all error levels, the best
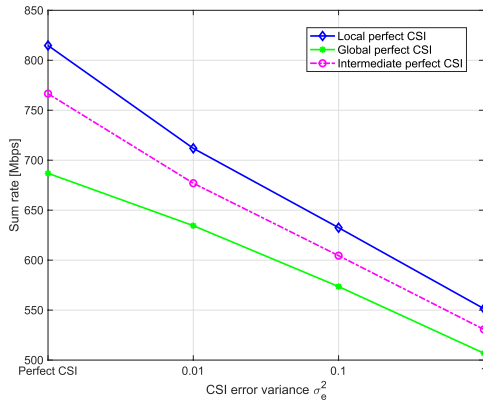
Fig. 8 Average sum-rate comparison of reference and proposed algorithms, different levels of CSIs at FogAPs, $T = 5$, $K = 90$.
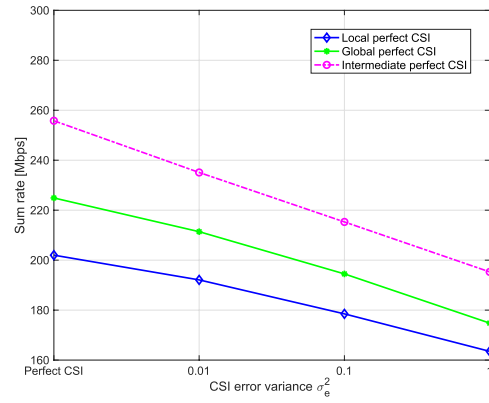

Fig. 10 Pico FogAPs' average sum-rate comparison of reference and proposed algorithms, different levels of CSIs at FogAPs, $T = 5$, $K = 90$.
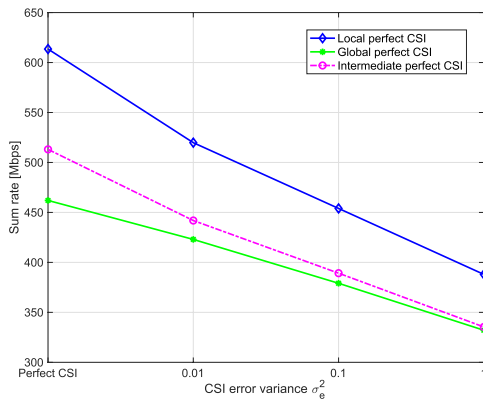

Fig. 9 Macro FogAPs' average sum-rate comparison of reference and proposed algorithms, different levels of CSIs at FogAPs, $T = 5$, $K = 90$.

posed scheme allows to improve the system throughput, energy efficiency and delays for large and small messages simultaneously, in the range of realistic global CSI imperfectness. In addition, the clustering optimization period T and perfect CSI knowledge level at each FogAP may be adapted given the required levels of computation complexities, allowed amounts of CSI feedback overhead and desired Pico-Macro FogAPs' performance trade-offs.

6. Conclusion

We have investigated the issues of optimized radio resource allocation in FogRANs under the practical assumption of imperfect CSI knowledge at cloud BBUs. We proposed a hybrid semi-distributed resource allocation algorithm with centralized user pre-scheduling carried out periodically at cloud BBUs, and distributed local beamforming at each FogAP in each frame. Although global, the centralized algorithm that jointly solves the user clustering and beamforming in CRANs can only make use of imperfect CSI due to the inevitable transport delays on fronthaul links. Therefore, our algorithm takes advantage of both the large-scale cloud processing to optimize the user pre-scheduling despite imperfect global CSI, and the availability of perfect but local CSI at FogAPs for accurate beamforming. Simulation results showed the effectiveness of the proposed method for realistic CSI qualities, as well as high robustness against CSI uncertainties. In particular, the delay improvements for short messages suggest that our approach is well-suited to support future IoT applications that typically generate a large amount of very small packets. Overall, the proposed design ensures an essential trade-off between the network-wide sum-rate and energy efficiency performance, required computational complexity at FogAPs and latency.

This work has identified key issues to be investigated, such as the optimized design of pre-scheduling and beamforming and CSI acquisition under high user mobility, as well as the learning and prediction-based joint edge caching and resource allocation issues in dynamic environments.

performance is achieved by local perfect CSI knowledge, while the global perfect CSI case performs worst. This is because with global perfect CSI, each FogAP has to minimize its own interference leakage towards all users in the cell, at the detriment of its own signal power, as can be seen in Eq. (11). The same trend can be observed with the average sum-rate achieved by Macro FogAPs in Fig. 9, but with a higher gain achieved by local perfect CSI case. However, this tendency is reversed concerning the average sum-rate of Pico FogAPs as shown in Fig. 10, where the best performance is given by intermediate perfect CSI knowledge. This can be understood as follows: given the much higher transmit powers of Macro FogAPs, Pico FogAPs greatly benefit from the inter-cell interference leakage reduction policies offered by intermediate and global CSI knowledge cases. While global CSI knowledge requires also the Pico FogAPs to decrease their leakage towards all users, intermediate CSI knowledge allows Pico FogAPs to care only about their own intra-cell interferences, thereby further improving the achievable sum-rate. Overall, intermediate perfect CSI provides the best balance between Macro and Pico FogAP sum-rates, while decreasing the amount of CSI feedback overhead as compared to the global perfect CSI case.

Observing all figures, we can conclude that the pro-

References

[1] N. Pontois, M. Kaneko, T.H.L. Dinh, and L. Boukhatem, "User pre-scheduling and beamforming with outdated CSI in 5G fog radio access networks," IEEE Globecom, Dec. 2018.

[2] A. Checko, H.L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M.S. Berger, and L. Dittmann, "Cloud RAN for mobile networks — A technology overview," IEEE Commun. Surveys Tuts., vol.17, no.1, pp.405–426, Sept. 2014.

[3] H. Dahrouj, A. Douik, O. Dhifallah, T.Y. Al-Naffouri, and M.S. Alouini, "Resource allocation in heterogeneous cloud radio access networks: Advances and challenges," IEEE Wireless Commun., vol.22, no.3, pp.66–73, Jun. 2015.

[4] L. Liu, S. Bi, and R. Zhang, "Joint power control and fronthaul rate allocation for throughput maximization in OFDMA-based cloud radio access network," IEEE Trans. Commun., vol.63, no.11, pp.4097–4110, Nov. 2015.

[5] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," IEEE Access, vol.2, pp.1326–1339, Oct. 2014.

[6] B. Dai and W. Yu, "Backhaul-aware multicell beamforming for downlink cloud radio access network," IEEE ICC Workshops (ICCW), Sept. 2015.

[7] Y. Shi, J. Zhang, and K.B. Letaief, "Group sparse beamforming for green cloud-RAN," IEEE Trans. Wireless Commun., vol.13, no.5, pp.2809–2823, May 2014.

[8] X. Huang, G. Xue, R. Yu, and S. Leng, "Joint scheduling and beamforming coordination in cloud radio access networks with QoS guarantees," IEEE Trans. Veh. Technol., vol.65, no.7, pp.5449–5460, July 2016.

[9] H.D. Thang, L. Boukhatem, M. Kaneko, and S. Martin, "Performance-cost trade-off of joint beamforming and user clustering in cloud radio access networks," IEEE PIMRC, Oct. 2017.

[10] H.D. Thang, L. Boukhatem, M. Kaneko, and S. Martin, "An advanced mobility-aware algorithm for joint beamforming and clustering in heterogeneous cloud radio access network," ACM MSWiM, Oct. 2018.

[11] D. Wang, Y. Wang, R. Sun, and X. Zhang, "Robust CRAN precoder design for wireless fronthaul with imperfect channel state information," IEEE WCNC, March 2017.

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

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

[14] J.H. Park, O. Simeone, and S. Shamai, "Joint optimization of cloud and edge processing for fog radio access networks," IEEE Trans. Wireless Commun., vol.15, no.11, pp.7621–7632, Nov. 2016.

[15] K. Hayashi, M. Kaneko, T. Fujii, H. Sakai, and Y. Okada, "Transmit beamforming and iterative water-filling based on SLNR for OFDMA systems," IEEE PIMRC, Sept. 2009.

[16] H. Du and P.J. Chung, "A probabilistic approach for robust leakage-based MU-MIMO downlink beamforming with imperfect channel state information," IEEE Trans. Wireless Commun., vol.11, no.3, pp.1239–1247, March 2012.

[17] M. Sadek, A. Tarighat, and A. Sayed, "Active antenna selection in multiuser MIMO communications," IEEE Trans. Signal Process., vol.55, no.4, pp.1498–1510, April 2007.

[18] R. Katsuki, M. Kaneko, and K. Hayashi, "A study on beamforming methods in CRAN with fronthaul link constraints," IEICE Technical Report, CQ2016-58, Aug. 2016.

[19] R.K. Mungara, G. George, and A. Lozano, "Overhead and spectral efficiency of pilot-assisted interference alignment in time-selective fading channels," IEEE Trans. Wireless Commun., vol.13, no.9, pp.4884–4895, Sept. 2014.

[20] J. Na, J. Koh, S. Park, and J. Kang, "Energy efficiency enhancement on cloud and edge processing by dynamic RRH selection," IEEE CCNC, Jan. 2018.

[21] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?," IEEE Trans. Wireless Commun., vol.14, no.6, pp.3059–3075, June 2015.

Megumi Kaneko received her MSc. degree in communication engineering (Diplôme d'Ingénieur) in 2004 from Télécom SudParis, France, jointly with a MSc. from Aalborg University, Denmark, where she received her Ph.D. degree in 2007. In May 2017, she obtained her HDR degree (French Doctoral Habilitation for Directing Researches at Professor position) from Paris-Sud University, France. From January to July 2007, she was a visiting researcher in Kyoto University, Japan, and a JSPS post-doctoral fellow from April 2008 to August 2010. From September 2010 to March 2016, she was an Assistant Professor in the Graduate School of Informatics, Kyoto University. She is currently an Associate Professor at the National Institute of Informatics as well as the Graduate University of Advanced Studies (Sokendai), Tokyo, Japan. Her research interests include wireless communications, radio resource and interference management, wireless signal processing and cross-layer network protocols. She received the 2009 Ericsson Young Scientist Award, the IEEE Globecom 2009 Best Paper Award, the 2011 Funai Young Researcher's Award, the WPMC 2011 Best Paper Award, the 2012 Telecom System Technology Award, the 2016 Inamori Foundation Research Grant and the 2019 Young Scientists' Prize from the Minister of Education, Culture, Sports, Science and Technology of Japan.

Lila Boukhatem is an associate professor at LRI laboratory, a joint CNRS and University of Paris Saclay/Paris-Sud 11 research Lab. She received a computer science engineering degree in 1997 from INI Institute (Algeria), MsC degree in 1998 from the University of Versailles Saint Quentin-en-Yvelines and Ph.D. degree in 2001 from the University of Pierre et Marie Curie (Paris 6). She joined the LRI laboratory in 2002 where she is developing various research works in mobile and wireless networks including: Cross-layer design, modeling and performance evaluation, radio resource allocation and optimization in 4G and 5G networks, mobility management, and routing and energy in ad hoc networks. She has been involved in several national and international projects and programs such as PHC Sakura, FP7 - NoE (Network of Excellence) NEWCOM#, and CNRS PICS. Her current research interests are interference mitigation in Heterogeneous Networks and resource allocation in Cloud Radio Access Networks (C-RANs).

Nicolas Pontois received his MSc. degree in communication engineering (Diplôme d'Ingénieur) in 2017 from Télécom Paris, France and from Université Paris-Sud, France. From March to September 2017, he was an internship student at the National Institute of Informatics, Tokyo, Japan, where he conducted part of this research work within the framework of the MoU grant between NII and Université Paris-Sud.

Thi Hà Ly Dinh received her BSc. and MSc. degree in computer science in 2015 and 2017 from Hanoi University of Science and Technology, Vietnam. She is currently working towards the Ph.D degree at the Graduate University of Advanced Studies (Sokendai), National Institute of Informatics, Tokyo, Japan. Her research interests include optimization problems, evolutionary algorithms, wireless communications, radio resource and interference management.