A Dynamic Transmission Strategy Based on Network Slicing for Cloud Radio Access Networks

Mahdi Ezzaouia, Cédric Gueguen, Melhem El Helou, Mahmoud Ammar, Xavier Lagrange, Ammar Bouallegue

► To cite this version:

Mahdi Ezzaouia, Cédric Gueguen, Melhem El Helou, Mahmoud Ammar, Xavier Lagrange, et al.. A Dynamic Transmission Strategy Based on Network Slicing for Cloud Radio Access Networks. 10th Wireless Days Conference, Apr 2018, Dubai, United Arab Emirates. hal-01972777
A Dynamic Transmission Strategy Based on Network Slicing for Cloud Radio Access Networks

Mahdi Ezzaouia*,†, Cédric Gueguen†, Melhem El Helou‡, Mahmoud Ammar†, Xavier Lagrange*, Ammar Bouallegue†

*IMT Atlantique, IRISA, 2 Rue de la Chataigneraie, 35576 Cesson Sévigné, France
Email: mahdi.ezzaouia@imt-atlantique.fr
†University of Tunis El Manar, National Engineering School of Tunis, Communications Systems Laboratory, 1002 Tunis, Tunisia
‡University of Rennes 1, IRISA, Campus de Beaulieu, 35042 Rennes, France
§Ecole Supérieure d’Ingénieurs de Beyrouth, Saint Joseph University of Beirut, Beirut, Lebanon

Abstract—The Cloud Radio Access Network (C-RAN) has clearly emerged as a promising evolution of wireless networks. This architecture consists in decoupling the baseband units (BBUs) from the remote radio heads (RRHs). The BBUs are pooled in the same centralized BBU pool, while the RRHs are distributed through different distant sites. Typically, the one-to-one logical mapping consists in assigning one BBU to one RRH so that distinct frames are generated for each RRH. Also, a logical mapping of one BBU to many RRHs could be established in order to deliver the same frames of a single BBU to a cluster of RRHs. Motivated by the network slicing concept, we propose a hybrid transmission strategy where the resource units (RUs) of each frame are partitioned in two slices. The first one constitutes the unshared slice and is allocated to the cell center users (CCUs) according to the one-to-one logical mapping. The second slice is constituted by a quantity of RUs shared by a cluster of RRHs that belong to the same BBU. This last common slice is transmitted according to the one-to-many mapping in order to be allocated to the cell edge users (CEUs) and to the mobile users (MUs). We also present a flexible solution that dynamically adjusts the BBU-RRH mapping scheme (i.e., one-to-one, one-to-many or hybrid) based on the radio resource usage of BBUs. On the one hand, our proposed solution achieves close packet delay to the one-to-one configuration, while providing lower power consumption and handover frequency. On the other hand, in comparison with the one-to-many configuration, our technique provides lower packet delay and subsequently better Quality of Service (QoS).

Index Terms—Cloud-RAN; Wireless networks; Resource allocation; Scheduling.

I. INTRODUCTION

The expectation of getting faster wireless internet connection speeds and higher system’s capacity becomes more challenging than ever. In this context, the capacity requirements of the future cellular networks can be supported via the dense deployment of base stations (BSs) with a high user equipments (UEs) activity. This massive deployment of nodes significantly increases the network capital and operating costs. In this context, the Cloud Radio Access Networks (C-RANs) [1] is a recent mobile architecture that brings potential solutions to these issues.

Traditionally, the C-RAN architecture is constituted by the remote radio heads (RRHs) and the baseband units (BBUs). The RRHs include radio antennas with their associated ampli-
ing problem as a bin packing problem. They further introduced an optimal and a heuristic solutions to reduce the number of active BBUUs, thus decreasing the power consumption, without compromising user QoS. Only adjacent RRHs are allowed to form a cluster, so as to reduce the handover frequency. The RRH clustering was expressed in [3] as a coalition formation game: disjoint clusters are formed in a way to maximize a network utility function. This function reflects the network performance in terms of throughput, power consumption and handover frequency. In this context, a centralized approach and a distributed technique, based on the merge-and-split rule, are introduced. Authors showed that their distributed solution reaches close performance to the centralized one, while reducing the number of iterations needed for the clustering process. However, the complexity of these proposed clustering algorithms is still high. Moreover, the majority of the RRH-BBUUs mapping approaches [2-5], proposed in the literature, are not adapted to realistic scenarios, where users are characterized with dynamic and variable traffic load requirements. As a matter of fact, they formulate the RRH clustering strategies considering a full buffer traffic model, that does not consider the traffic dynamicity.

In this paper, we propose a novel and intelligent technique based on a logically re-configurable front-haul to enhance the performances of the C-RAN systems. Our main contributions are summarized as follows:

- We introduce a novel approach that consists of dividing the radio resources of each frame into two different slices. The first one is allocated to the CCUs, according to the one-to-one mapping, in order to increase the provided throughput. As a matter of fact, this slice is reused in the cell center area of each RRH since the CCUs have a good signal quality. The second slice is allocated to the CEUs and MUs following the one-to-many approach, so that the inter-cell interferences and the handover frequency are reduced. Hence, the performances of the CEUs and the MUs are enhanced.
- We propose an intelligent algorithm implemented in the BBU pool in order to determine the best transmission strategies (i.e., one-to-one, one-to-many or hybrid) based on the global radio resource usage. In other words, our proposal adapts to network load conditions, taking advantages of the one-to-many mapping at low load conditions and of the one-to-one mapping at high load conditions.

The rest of this article is organized as follows: In section II, we introduce and explain our proposed solution. The system model and the performance evaluation are presented in section III. Section IV concludes this paper.

II. THE PROPOSED SOLUTION

A. The C-RAN model

The transmission on the radio interface is based on the Orthogonal Frequency Division Multiplexing (OFDM), and the bandwidth is split into \( N \) subcarriers or sub-frequency bands. In the time domain, the frequency resources are divided in frames which are constituted by time slots having the same constant duration. Each pair of subcarrier and time slot constitutes an elementary resource unit (RU). Our C-RAN system is constituted by \( R \) RRHs (cells) connected to the BBU pool through high-performance optical front-haul links. We also consider \( K \) UEs distributed through the network. In the downlink, the packets generated from the core network are stored in the buffers of the BBU pool which forward them to users through the RRHs. We suppose that the Dynamic Transmission Point Selection [6] technique is implemented at the BBU pool, so that each UE can receive data from only one RRH at a time. The scheduler is located at the MAC layer of the BBU and manages the buffers in a centralized manner. In this work, we choose the Proportional Fair (PF) scheduler that consists in allocating the RU to a mobile \( j \) when its channel conditions are the most favorable with respect to its time average:

\[
    j = \arg\max_k \left( \frac{m_{k,n}}{D_k} \right), \quad k = 1, \ldots, K, \tag{1}
\]

where \( m_{k,n} \) is the maximum number of bits that can be transmitted over RU \( n \) if allocated to mobile \( k \). This value is computed as a function of user \( k \) radio conditions as detailed in subsection III-A. The parameter \( D_k \) is the average throughput provided by the scheduler to user \( k \) during the last scheduling periods [7].

The frame generated by each BBU is composed by \( \frac{N_{RU}}{N} \) resource units. In the case of a one-to-one configuration, each
RRH is associated to a BBU. Users in the vicinity of each RRH are thus scheduled in different frames. In this context, all the bandwidth is allocated to each RRH, and the spectrum is reused in each cell. For the one-to-many mapping, all users in the different RRHs share the same RUs of an only one BBU: they are scheduled and multiplexed within the same frame.

B. The network slicing-based approach

On the one hand, the one-to-one logical mapping is best suited for static UEs with high traffic load profiles. On the other hand, the one-to-many configuration is more adapted for lightly loaded scenarios and efficiently supports user mobility. In order to benefit from the advantages of these two kinds of transmission strategies, we introduce a hybrid configuration mode. Our solution is motivated by the network slicing concept \cite{8}. In fact, we assume that the RUs of each frame are divided into two slices. Also, UEs are classified into two groups: the first one contains the CCUs, and the second one is constituted by the CEUs and the MUs. On the one hand, the first slice of the frame is exclusively dedicated to CCUs within the cluster. This slice of frame is generated separately by each BBU associated to each RRH and is transmitted according to the one-to-one mapping. This allows to increase the provided throughput, particularly at high traffic load conditions. On the other hand, the second slice of each frame is composed by a number of RUs to be assigned to all the CEUs and the MUs of the cluster. This number of RUs constitutes a common part of the frame and is generated to the RRHs according to the one-to-many mapping so as to reduce the inter-cell interference. Also, the solution better supports user mobility. For illustration, in Fig. 2 we represent the hybrid frame structure, where users are represented by different colors. As the Dynamic Transmission Point Selection technique is implemented in the BBU pool, each RRH transmits the RUs of only its associated users. That is why, in the shared slice of a given RRH, the RUs allocated to users that belong to another RRH are locked (represented by a hatched area).

The RUs repartition between the two slices highly depends on the number of UEs in the two groups. In our work, we assume, that for an OFDMA frame, RUs are divided as follows:

$$N_{RUsh} = \left\lfloor \left( \frac{K_e + K_m}{K} \right) N_{RU} \right\rfloor,$$

$$N_{RUunsh} = N_{RU} - N_{RUsh},$$

where $N_{RU}$ is the total number of RUs in a frame generated by a single BBU, $N_{RUsh}$ represents the number of shared RU, $N_{RUunsh}$ is the number of unshared RU, $K_e$ is the number of CEUs and $K_m$ is the number of MUs.

In the hybrid frame configuration, CCUs suffer from higher level of interferences in comparison with the one-to-many configuration. In this context, it would be clear that, at a low traffic load, the one-to-many system provides better QoS for the CCUs than the hybrid configuration. Furthermore, in the one-to-one mapping, the spectrum is reused by each BBU so that it generates a higher level of interferences mainly for the CEUs. In this case, the hybrid solution performs better than the one-to-one configuration at low traffic load. Moreover, the one-to-one mapping consumes higher energy than the hybrid and the one-to-many configurations, as more RUs are used and more BBUs are active. This leads to low energy efficiency. However, at extremely high traffic load and with static users, the one-to-one clustering scheme provides better QoS than the two others solutions, at the cost of a higher power consumption. In fact, more RUs are available, and consequently the C-RAN system provides higher throughput.

C. The dynamic transmission strategy

In order to adapt our solution to the non-uniform traffic load distribution and to the UEs mobility while reducing the power consumption, we propose to implement an intelligent controller at the BBU pool. The objective is to dynamically reconfigure the BBU-RRH mapping scheme. In fact, this entity consists in fixing the appropriate transmission strategy (i.e., one-to-many, hybrid and one-to-one) that satisfy user throughput requirements in a given cluster. Thus, it employs an efficient and dynamic algorithm to determine the best transmission strategy which provides a trade-off between the provided QoS, the power consumption and the handover frequency.

The proposed solution is given by Algorithm 1 where $RW_{used}$ is the radio resource usage ratio of BBUs and is
computed by the scheduler of the BBU pool. This parameter depends on the current BBU-RRH mapping mode:

- In case of a one-to-one mapping, each active BBU computes its own radio resource usage ratio defined as the total number of RU allocated by the scheduler divided by the total number of RU. We then consider the smallest value among them.
- In case of a one-to-many configuration, we consider the radio resource usage ratio of a single BBU mapped to the considered cluster of RRHs.
- In case of a hybrid configuration, it corresponds to the radio resource usage ratio of the shared slice (Fig. 2). In fact, we only consider the ratio of used RU divided by the total number of RUs in the shared slice.

Finally, the threshold parameters \( RW_1 \) and \( RW_2 \) are tuned to optimize our dynamic algorithm, since they directly reflect the traffic load of the network. A high \( RW_1 \) value leads to more resource sharing between the RRHs and subsequently to lower power consumption. Yet, a low \( RW_2 \) leads to more resource reuse, but also to high interference typically at the cell edge.

**Algorithm 1** The BBU-RRH Mapping Algorithm

1. BEGIN
2. Update \( (RW_{\text{used}}) \)
3. if \( (RW_{\text{used}} \leq RW_1) \) then
4. Set the one-to-many mapping as a transmission strategy
5. else if \( (RW_1 < RW_{\text{used}} \leq RW_2) \) then
6. Set the hybrid mode as a transmission strategy
7. else
8. Set the one-to-one mapping as a transmission strategy
9. end if
10. Schedule the UEs based on the selected strategy.
11. END

### III. PERFORMANCE EVALUATION

#### A. The channel model

The RRHs are equipped with an omni-directional antenna installed in the center of each hexagonal cell. The channel gain \( G^i_{k,n} \), between the serving RRH \( i \) and user \( k \) on subcarrier \( n \) is given by:

\[
G^i_{k,n} = h \times 10^{\frac{X \sigma}{10}} \times \left( \frac{d_{0}}{d_{k,i}} \right)^{\alpha}.
\]

\( h \) represents the Rayleigh multipath fading, which is modeled by an exponential distribution, \( X \) is a standard gaussian random variable, \( \sigma \) is the standard deviation of shadowing in dB, \( d_{k,i} \) is the distance between UE \( k \) and RRH \( i \), \( d_{0} \) is the reference distance and \( \alpha \) is the path loss exponent.

We assume that RRH \( i \) belongs to a cluster formed by a set \( C \) of RRHs mapped to the same BBU. We further denote by \( C' \) the set of \( L \) RRHs which are adjacent to the cluster \( C \) \((C \cap C' = \emptyset)\). In this context, all the interfering signals in RRH \( i \) originate from the RRHs that belong to the cluster \( C' \), as they are mapped to different BBU. Therefore, the signal to interference plus noise ratio of user \( k \) on subcarrier \( n \) associated to RRH \( i \) is given by:

\[
\gamma^i_{k,n} = \frac{P^i_n G^i_{k,n}}{B_{\text{sub}} N_0 + \sum_{j=1, j \in C'} P^j_n G^j_{k,n}},
\]

where \( P^i_n \) and \( P^j_n \) are respectively the transmitted power on subcarrier \( n \) of RRH \( i \) and the transmitted power on subcarrier \( n \) of the interfering RRH \( j \) that belongs to the cluster \( C' \). Also, \( G^i_{k,n} \) is the channel gain between UE \( k \) and RRH \( i \), the parameter \( N_0 \) is the thermal noise power density, and \( B_{\text{sub}} \) is the subcarrier spacing [7].

We used the adaptive Modulation and Coding (AMC) in order to adapt the modulation and coding scheme to the quality of the received signal. We followed the procedure described in [9] to perform the AMC. We start by using Shannon’s formula to compute the spectral efficiency \( \eta^i_{k,n} \) of UE \( k \) on subcarrier \( n \) associated to RRH \( i \) as follows:

\[
\eta^i_{k,n} = \log_2 \left( 1 + \frac{\gamma^i_{k,n}}{\Gamma} \right),
\]

where \( \Gamma = -\ln(5E)/1.5 \) is a SNR correction factor that takes into account the difference between the information-theoretic performances and the practical implementation of the MCS [10], and \( E \) is a BER Target. Finally, we use the LTE CQI efficiency table [9] to determine the number of bits that could be transmitted to a UE over a subcarrier.

#### B. Simulation scenario

For illustration, we consider a small network constituted by two adjacent hexagonal cells surrounded by eight neighboring ones. In the one-to-many mapping, we suppose that the two considered RRHs collaborate together in order to form independent clusters. The UEs are evenly distributed into three groups according to their profiles: the first group contains the CCUs whose distance from their respective RRH is less than the interior radius of the cell. The second group is constituted by the CEUs, and the last one represents the MUs that are assumed to move at a constant speed. Moreover, the MUs have a direction perpendicular to the straight line that separates the two central hexagonal cells (Fig. 1). We also suppose that they are going back and forth between the extremities of the cluster.

All the clients run the same type of real time application that produces high peak bit rates with high burstiness and tight delay constraints. This greatly complicates the task of packet schedulers. In fact, studying the performance of resource allocation strategies with real time traffic totally leverages the performances of the system compared to the full buffer model (where the buffer are assumed to be always full). Also, we assume that each UE has only one service flow with a traffic composed of an MPEG-4 video stream [11], which is a very realistic and complex kind of traffic (video-conference). In addition, it creates a high volume of data with high sporadicity. These high peak bit rates greatly affect the performance of the system by abruptly filling the buffers during brief periods of time [9]. The average bit rate of each UE is set to 150 kbps.
Transmission strategies for 30 UEs.

Transmission strategies for 48 UEs.

Transmission strategies

One-to-many

One-to-one

Hybrid

The proposed solution

Figure 3: Time variation of the proposed solution between the different transmission strategies.

In a first simulation, each RRH of the two target adjacent cells contains 6 UEs uniformly distributed among the three groups. We then successively increase the number of UEs per group in the two RRHs. Therefore, in each simulation run, six UEs are added to the network (3 UEs per cell). Finally, the threshold parameter $RW_1$ and $RW_2$ are respectively fixed to 0.7 and 0.9. In this paper, they are considered as static parameters, but in future work, they can be optimally determined to meet operator objectives. The rest of the simulation parameters are listed in Table 1.

C. Performance metrics

1) Packet delay: We assume that UEs’ traffic streams are arranged in blocks of bits, denoted as packets, having the same constant size at the MAC level. The packet delay is the time between the packet arrival in the transmission buffer and its reception time at the UE. It corresponds to the packet waiting time in the service flow transmission buffer, if the transmission and propagation delays are neglected.

2) Power Consumption: The power consumed by BBUs associated to a cluster of RRHs is a linear function of the number of RUs used by each BBU. The power consumption is given as follows:

$$PC = \lambda + \mu.RU(b,n),$$

where $\lambda$ is the minimum amount of power consumed by an active BBU (at 0 load), $\mu$ represents the variation coefficient of the power consumption as a function of RU(b,n). The parameter RU(b,n) captures the BBU’s resource usage and is expressed as:

$$RU(b,n) = N_{RUsh}.1 + (N_{RU} - N_{RUsh}).n.$$  

Table I: Simulation parameters.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of RRHs</td>
<td>10</td>
</tr>
<tr>
<td>Cell Radius ($R$)</td>
<td>500 m</td>
</tr>
<tr>
<td>Interior radius</td>
<td>$2R/\beta$</td>
</tr>
<tr>
<td>Number of subcarriers</td>
<td>75</td>
</tr>
<tr>
<td>RRH transmit power</td>
<td>20 W (43 dBm)</td>
</tr>
<tr>
<td>Standard deviation of shadowing</td>
<td>$\sigma = 8$dB</td>
</tr>
<tr>
<td>Reference Distance ($d_0$)</td>
<td>1 km</td>
</tr>
<tr>
<td>Path-loss exponent</td>
<td>3.5 (urban area)</td>
</tr>
<tr>
<td>Target BER</td>
<td>$5 \times 10^{-4}$</td>
</tr>
<tr>
<td>Subcarrier spacing</td>
<td>15 kHz</td>
</tr>
<tr>
<td>Thermal noise power density $(N_0)$</td>
<td>$-174$ dBm/Hz</td>
</tr>
<tr>
<td>$RW_1$</td>
<td>0.7</td>
</tr>
<tr>
<td>$RW_2$</td>
<td>0.9</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>50 W</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.6</td>
</tr>
</tbody>
</table>

3) Handover: The handovers highly overload the network in terms of signaling messages while reducing the system’s performances. Thus, it is important to consider this parameter in our simulation model. In this context, we assume that handovers are only triggered based on the distance between the MU and its current associated RRH: the mobile is associated to its nearest RRH. We adopt the global number of handovers as a metric. In fact, this parameter corresponds to the number of times a UE crosses the border between the two target RRHs when the one-to-one transmission strategy is active.

D. Results

Fig. 4 shows the variation of the transmission strategies as a function of time for an underloaded C-RAN system with 12 UEs (Fig. 3(a)), a moderate load system with 30 UEs (Fig. 3(b)) and for a highly loaded one with 48 UEs (Fig. 3(c)). As we can see, the proposed solution employs the one-to-many mapping at low user concentration in order to reduce the power consumption since half of the BBUs are used. At moderate user concentration, our proposal switches between the hybrid and the one-to-one configuration, since UEs are characterized by a dynamic and variable traffic. Finally, at high user concentration, our technique configures the transmission scheme as the one-to-one logical mapping to effectively cope with user needs in terms of radio resources.

In Fig. 4(a), we represent the mean packet delay of users in the two target adjacent cells as a function of the number of users. For simplicity, in this work we do not consider the impact of handovers while computing the delay of the MUs. Consequently, the results we present are optimistic for the one-to-one mapping. We notice that all the solutions provide approximately the same results in terms of delay when the network is underloaded. However, we note that the one-
to-many configuration provides the worst results when the number of users increases. In fact, all the UEs are multiplexed within the same frames, and the number of available RUs becomes insufficient to serve all the UEs. However, our proposed technique reaches the performance results of the one-to-one mapping regardless the number of users.

The Fig. 4(b) shows the mean power consumed by BBUs of the two target RRHs as a function of users concentration. The one-to-one mapping consists to assign one BBU to each RRH and consumes more RUs than the one-to-many mapping. Consequently, it reaches always the same constant level of energy consumption regardless of the number of users in the system. The solution we propose aims to reduce the power consumption of the C-RAN system while providing an acceptable QoS. In this context, when the number of users is low, it provides the same results as the one-to-many configuration and outperforms the one-to-one mapping. At moderate traffic load, our solution employs more frequently the hybrid frame structure (Fig. 3(b)) and provides a trade-off between the power consumption reduction (Fig. 4(b)) and the provided QoS (Fig. 4(a)).

Fig. 4(c) shows the mean number of handovers in the target cluster. As we use a discrete event simulator, the absolute results highly depend on the user mobility model. The one-to-one mapping reaches always the worst results, since the RRHs are mapped to different BBUs. In the one-to-many mapping, as MUs within the same cluster belong always to the same BBU, no handover is established. Our proposed scalable technique determines and adjusts the transmission strategy based on the resource usage. Consequently, the handovers occur only when the one-to-one mapping mode is active.

IV. CONCLUSION

In this paper, we have investigated different BBU-RRH mapping strategies for C-RAN architectures. We introduced a novel hybrid frame configuration model, combined with a dynamic algorithm, in order to determine and apply the appropriate transmission strategy. Simulation results show that the proposed solution adapts to network load conditions, reaching the performances of the one-to-many mapping at low load conditions and of the one-to-one mapping at high load conditions. In the near future, we will study the importance of $R_W^1$ and $R_W^2$ to the system performances. In this context, they can be tuned, based on an optimization algorithm, to meet operator objectives and strategies. Finally, we would like to underline that this research study motivates future work on transmission techniques using logical re-configurable front-hauls in order to adapt the C-RAN architecture to the non-uniform user profiles with various traffic load conditions.

REFERENCES