Resource provisioning in wireless virtualized networks via massive-MIMO

This item was submitted to Loughborough University's Institutional Repository by the/an author.


Additional Information:

• (c) 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.

Metadata Record: [https://dspace.lboro.ac.uk/2134/21229](https://dspace.lboro.ac.uk/2134/21229)

Version: Accepted for publication

Publisher: © IEEE

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: [https://creativecommons.org/licenses/by-nc-nd/4.0/](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Please cite the published version.
Resource Provisioning in Wireless Virtualized Networks via Massive-MIMO

Vikas Jumba*, Saeedeh Parsaeefard*, Mahsa Derakhshani†, Tho Le-Ngoc*
*Department of Electrical & Computer Engineering, McGill University, Montreal, QC, Canada
†Department of Electrical & Computer Engineering, University of Toronto, Toronto, ON, Canada

Email: vikas.jumba@mail.mcgill.ca; saeedeh.parsaeifard@mcgill.ca; mahsa.derakhshani@utoronto.ca; tho.le-ngoc@mcgill.ca

Abstract—This letter proposes a dynamic resource provisioning scheme for an OFDMA wireless virtualized network (WVN), where one base-station equipped with a large number of antennas serves users belonging to a number of service providers via different slices. In particular, joint power, sub-carrier, and antenna allocation problems are presented for both perfect and imperfect channel knowledge cases, aiming to maximize a sum-utility while maintaining a minimum rate per slice. Subsequently, relaxation and variable transformation are applied to develop the efficient algorithm to solve the formulated non-convex, combinatorial optimization problem. Simulation results reveal the benefits of applying a large number of antennas in this setup and evaluate the network performance for different system conditions.

Index Terms—Large number of antennas, resource provisioning, wireless network virtualization.

I. INTRODUCTION

Wireless network virtualization is a promising approach to provide service customization for next-generation wireless networks. In wireless virtualized networks (WVN), limited wireless resources (i.e., power and spectrum) are shared among different groups of users referred to as slices. Each slice may have a different set of quality-of-service (QoS) requirements, which calls for effective resource provisioning algorithms in WVN to maximize the sum utility of all slices, while holding the minimum required rate of each slice [1].

However, due to the random nature of wireless channels and power and spectrum limitations, there is always a non-zero probability that the minimum required rate cannot be satisfied. Thus, this will cause infeasibility problem for resource provisioning and lead to a poor network performance. In this paper, to expand the feasibility region, we take advantage of the large spatial degree of freedom introduced by massive multi-input multi-output (MIMO) technology. In massive MIMO, each base station (BS) is equipped with a large number of antennas serving single-antenna multiple users.

Developing the massive-MIMO scales up the MIMO multiplexing gain provided that the exact channel state information (CSI) of all users are available at BS [2]. However, extracting the precise values of CSI requires ideal orthogonal pilot signals between different BSs which is practically infeasible [2]. In this paper, taking this issue into account, we consider two cases: 1) perfect CSI where it is ideally assumed to have all the precise values of CSI; 2) imperfect CSI where the effect of non-orthogonal pilots is considered in the achievable rate.

Subsequently, we formulate a resource provisioning problem for the up-link transmission in a WVN. The objective is to maximize the sum utility of slices subject to the minimum required rate of each slice and transmit power of each user within each slice. We introduce a new utility function representing the difference between the total achieved user rate and its corresponding costs for allocated power and antennas. Considering two different pricing mechanisms for the allocated power and antennas enables effective control of inter-slice interference and available antennas in WVN. To the best of our knowledge, there exists no related work in the context of WVN with massive MIMO.

As the formulated problem is both non-convex and combinatorial, it suffers from high computational complexity. To develop an efficient algorithm, we apply variable transformations and constraint relaxations. Simulation results reveal that the values of power and antenna pricing variables have a significant impact on the total achieved user throughput.

The rest of this paper is organized as follows. Section II introduces the system model and problem formulations. Section III provides a solution for the formulated resource provisioning problem followed by Section V where the simulation results are presented. Section VI provides concluding remarks.

II. NETWORK MODEL AND PROBLEM FORMULATION

We consider the up-link transmission in an orthogonal frequency division multiple access (OFDMA) WVN where a BS with $M$ antennas serves a set of slices $G = \{1, \ldots, G\}$. Each slice $g \in G$ has a set of single-antenna users denoted by $N_g = \{1, \ldots, N_g\}$ and requires a minimum rate $R_{g}^\text{min}$. It should be noted that $N = \sum_{g \in G} N_g$ and $N \ll M$. Consider $\mathbf{M} = [\mathbf{M}_1, \ldots, \mathbf{M}_G]$ as the allocated antenna vector for all slices where $\mathbf{M}_g = [\mathbf{M}_{ng,1}, \ldots, \mathbf{M}_{ng,K}]$, and $M_{ng,k}$ is the number of antennas allocated to user $n_g$ on sub-carrier $k$. Let $\mathbf{w} = [\mathbf{w}_1, \ldots, \mathbf{w}_G]$ be the sub-carrier assignment vector for all slices where $\mathbf{w}_g = [\mathbf{w}_{ng,1}, \ldots, \mathbf{w}_{ng,K}]$, and $w_{ng,k} = 1$ indicates that sub-carrier $k$ is allocated to user $n_g$ and otherwise $w_{ng,k} = 0$.

In this setup, let $\mathbf{h}_g, \mathbf{b}_g \in \mathbb{C}^{1 \times M_{ng,k}}$ be the channel vector of user $n_g$ on sub-carrier $k$, where $\mathbf{h}_{ng,k,m_{ng,k}}$ is the channel coefficient of user $n_g$ on sub-carrier $k$ and antenna $m_{ng,k}$. More specifically, $\mathbf{h}_{ng,k,m_{ng,k}} = h_{ng,k,m_{ng,k}} \sqrt{d_{ng,k}}$, where $h_{ng,k,m_{ng,k}}$ represents the small-scale fading coefficient with variance of 1, and $d_{ng,k}$ denotes the large-scale fading coefficient of user $n_g$ on sub-carrier $k$. Note that $d_{ng,k}$ includes both path loss and shadowing [2]. Practically, the channel
coefficients are estimated by the BS based on the uplink pilots with duration $\tau$ at the specific part of the coherence interval of $T$ [3]. In the perfect CSI case, the up-link received sample at BS after using the linear detector from user $n_g \in N_g$ on sub-carrier $k$ is [3]

$$y_{n_g,k}^{\text{perf}} = \sqrt{P_{n_g,k}} h_{n_g,k}^{\text{perf}} x_{n_g,k} + I_{n_g,k} + z_{n_g,k}^{\text{perf}},$$

where $x_{n_g,k}$ and $h_{n_g,k}^{\text{perf}} \in \mathbb{C}^{M_{n_g,k} \times 1}$ represent the transmit symbol and the precoding vector of user $n_g$ on sub-carrier $k$, respectively. Moreover, $P_{n_g,k}$ is the transmit power of user $n_g$ on sub-carrier $k$ and $z_{n_g,k}^{\text{perf}}$ is a vector of additive white Gaussian noise (AWGN) at the BS with zero mean and power spectral density $\sigma^2$, due to pilot contamination error in the channel estimation and the linear detector, the received signal from user $n_g$ on sub-carrier $k$ is [3]

$$y_{n_g,k} = \sqrt{P_{n_g,k}} h_{n_g,k} x_{n_g,k} + I_{n_g,k} + z_{n_g,k},$$

where $h_{n_g,k}$ is the estimated channel vector including the contamination error, $P = [P_1, \ldots, P_G]$ is the allocated power vector of all slices in which $P_g = [P_{n_g,1}, \ldots, P_{n_g,K}]$, $I_{n_g,k}^{\text{perf}} \in \mathbb{C}^{M_{n_g,k} \times 1}$ is the preceding vector in the imperfect CSI case, and $\delta_{n_g,k}(P)$ is the function of contamination error, interference and noise (See Appendix). Using maximum ratio combining (MRC) detector with $M \rightarrow \infty$, the rate of user $n_g$ on sub-carrier $k$ is derived in Appendix as

$$R_{n_g,k} = \begin{cases} \log_2 \left( 1 + P_{n_g,k} d_{n_g,k} M_{n_g,k} \right), & \text{for perfect CSI,} \\ \log_2 \left( 1 + \tau P_{n_g,k} d_{n_g,k}^2 M_{n_g,k} \right), & \text{for imperfect CSI,} \end{cases}$$

and the total achieved rate of user $n_g$ is $R_{n_g} = \sum_{k \in K} R_{n_g,k}$. Now, we define utility function of slice $g$ as

$$U_g(P, w, M) = \sum_{n_g \in N_g} R_{n_g} - c_g^M \sum_{n_g \in N_g} M_{n_g,k} - c_g^p \sum_{n_g \in N_g} P_{n_g,k},$$

where $c_g^M$ and $c_g^p$ are pricing factors for the number of allocated antennas and the transmit power of slice $g$, respectively. Hence, the resource provisioning problem can be written as

$$\max_{P, w, M} \sum_{g \in G} U_g(P, w, M),$$

subject to the following constraints C1 – C4.

C1: Exclusive sub-carrier allocation in OFDMA implies:

$$w_{n_g,k} \in \{0, 1\} \text{ and } \sum_{g \in G} \sum_{n_g \in N_g} w_{n_g,k} \leq 1, \forall k \in K.$$
Since (6) involves continuous variables and convex functions, we can solve it with the following Lagrange function.

\[
L(x, w, y, \lambda_n, \phi_g, \theta_g, \psi_g) = \sum_{g \in G} \tilde{R}_g + \sum_{n_g \in N_g} \lambda_n (\sum_{k \in K} x_{ng,k} - P_{\text{max}}) \\
+ \sum_{g \in G} \phi_g (R_{g,\text{av}} - \sum_{n_g \in N_g} \tilde{R}_{ng}) \\
+ \sum_{g \in G} \theta_g (M_{g,\text{min}} - \sum_{n_g \in N_g} \sum_{k \in K} y_{ng,k}) \\
+ \sum_{g \in G} \psi_g (\sum_{n_g \in N_g} \sum_{k \in K} y_{ng,k} - M_g^*),
\]

where \(\lambda_n, \phi_g, \theta_g\) and \(\psi_g\) are the Lagrange multipliers for \(C_2, C_3\) and \(C_4\). Now, we propose an iterative approach to solve the dual problem of (7), in the following algorithm.

Algorithm:

Initialization:
Set \(w^*(l = 0) = 1\), \(P_{n_g,k}(l = 0) = P_{\text{max}}/K\), and \(M^*(l = 0) = M_g^\text{max}^\text{}\) for all \(n_g \in N_g\) and \(g \in G\). Initialize \(l_{\text{max}} > 1\), \(0 < \varepsilon < 1\), \(\lambda(l = 0)\), \(\psi(l = 0)\) and \(\theta(l = 0)\).

Repeat:

- Update dual variables, \(\lambda_n, \phi_g, \theta_g\) and \(\psi_g\), by gradient descent method for all \(g \in G\) where \([x]^+ = \max\{x, 0\}\):

\[
\lambda_n(l + 1) = \left[\lambda_n(l) + \delta_n \frac{\partial L}{\partial \lambda_n} \right]^+, \forall n_g \in N_g, \\
\phi_g(l + 1) = \left[\phi_g(l) + \delta_g \frac{\partial L}{\partial \phi_g} \right]^+, \forall g \in G, \\
\theta_g(l + 1) = \left[\theta_g(l) + \delta_g \frac{\partial L}{\partial \theta_g} \right]^+, \forall g \in G, \\
\psi_g(l + 1) = \left[\psi_g(l) + \delta_g \frac{\partial L}{\partial \psi_g} \right]^+, \forall g \in G.
\]

- Using the above updated parameters for \((l + 1)\), compute \(P_{n_g,k}^*(l + 1)\) and \(M_{n_g,k}^*(l + 1)\) for all \(n_g \in N_g\) and \(k \in K\).

Finally, the optimum integer \(M_{n_g,k}^*\) is selected as \(\lceil M_{n_g,k}^* \rceil\) where \(\lfloor x \rfloor\) denotes the largest integer less than or equal to the value of \(x\).

IV. Simulation Results

To evaluate the performance of WVN with the proposed algorithm, we consider one BS servicing two slices (i.e., \(G = 2\)) with \(N_1 = N_2 = 4\). Furthermore, we assume \(K = 64\) and \(P_{\text{max}} = 0\) dB. After precoding, channel is modelled as \(|\mathbf{h}_{n_g,k,m_g,k}|^2 = d_{ng} = D_{ng}^{-\beta}\) where \(\beta = 3\) is the path loss exponent and \(D_{ng} \in [0.2, 0.6]\) is the distance of user \(n_g\) to BS. Furthermore, we set \(\varepsilon = 10^{-4}\) for Algorithm 1 and \(R_{g,\text{av}} = R_{g,\text{av}} = 2\) bps/Hz unless otherwise stated. \(T\) is selected based on the parameters in [3] and simulations are based on the ratio of \(T/T\). All the results are presented in terms of average rate over 100 channel realizations based on random locations of the users. When \(R_{g,\text{av}}\) does not hold for a channel realization, the total rate is set to zero. Also, to demonstrate the effect of \(c_g^M \) and \(c_g^P\), we consider three sets with different values of \(c_g^M \) and \(c_g^P\). For Set 1, we have \(c_g^M = c_g^P = 0\) which means that the utility of each user is equal to its rate. For Set 2, we have \(c_g^M = 0.07\) and \(c_g^P = 1\). Set 3 imposes higher restrictions to use antennas and power by employing \(c_g^M = 0.09\) and \(c_g^P = 2\).

Fig. 1 shows that the total rate is a non-decreasing function of \(M_g^\text{max}^\text{}\) in both perfect and imperfect CSI cases where \(\frac{T}{T} = 0.3\). This can be explained by the fact that the feasibility regions are expanded as \(M_g^\text{max}^\text{}\) increases. However, the utility function is formulated as a decreasing function of \(c_g^M \) and \(c_g^P\), so that by increasing the costs from Set 1 to Set 3, the total rate of WVN is decreased. As a result, the increase in the achieved rate is reduced for large values of \(M_g^\text{max}^\text{}\), especially for high \(c_g^P\) and/or \(c_g^M\), e.g., for \(M_g^\text{max}^\text{} > 75\), the total rate for Sets 2 and 3 is almost unchanged. Furthermore, due to the channel estimation errors, the performance with imperfect CSI is worse than that with perfect CSI for all sets, as expected.

To demonstrate the effects of \(R_{g,\text{av}}\), Fig. 2 plots the total rate versus \(R_{g,\text{av}}\) for Set 1, where \(\frac{T}{T} = 0.3\), \(M_g^\text{max}^\text{} = 200\), and \(R_{g,\text{av}} = R_{g,\text{av}} = 2\) bps/Hz. It can be observed that in both perfect and imperfect CSI cases, the total rate decreases with increasing \(R_{g,\text{av}}\). This is because increasing \(R_{g,\text{av}}\) shrinks the feasibility region for convex optimization problem (5) and thus reduces the optimal value of objective function [5].

To study further the effect of \(T\) on the performance of WVN, we focus on Set 1 with \(M_g^\text{max}^\text{} = 200\) and plot the rate versus \(\frac{T}{T}\) in Fig. 3. Obviously, when \(0.1 \leq \frac{T}{T} \leq 0.3\), the rate increases with increasing \(\frac{T}{T}\) benefiting from the more accurate CSI estimation. However, for \(0.3 \leq \frac{T}{T} \leq 0.9\), the spectral efficiency of WVN is decreased with increasing \(\frac{T}{T}\) due to the fact that most of the transmission time wastes for pilot signals. Therefore, considering the optimum values of \(\frac{T}{T}\) to reach the best performance of WVN is essential. Fig. 3 also highlights that increasing \(M_g^\text{max}^\text{}\) can increase the total rate for any \(T\) due to the feasibility region expansion.

To further study the effects of pricing on the total rate of WVN, in Fig. 4, the total rate is plotted versus \(c_g^M \) and \(c_g^P\). Obviously, the total rate is decreased with increasing \(c_g^M\) and \(c_g^P\).
Fig. 2: Total rate vs. $P^\text{Pilot}_g$

Fig. 3: Total rate vs. $\tau/T$

Fig. 4: Total Rate vs. $r_p^M$ and $c_p^M$

However, the effect of increasing $c_p^M$ is much more profound compared to that of $r_p^M$. This is because in massive MIMO, we have $M^\text{max}_g > P^\text{max}_n$, e.g., compare $M^\text{max}_g > 100$ and $P^\text{max}_n = 0$ dB. Figures 3 and 4 highlight that allocating dynamic and optimal pricing values for WVN and optimum value of $\tau/T$ are essential to reach the best performance which are left for the future work.

V. CONCLUSION

Utility-based resource provisioning for massive-MIMO-based WVN was investigated in this letter for perfect and imperfect CSI scenarios. We propose an efficient iterative algorithm to solve the developed resource allocation problems. Via simulation results, the effects of power and antenna pricing mechanisms on the performance of WVN were investigated.

VI. APPENDIX: DERIVATION OF USER RATE $R_{n_g,k}$ IN (3)

For mutually independent $1 \times n$ vectors $p = [p_1, \ldots, p_n]$ and $q = [q_1, \ldots, q_n]$ whose elements are i.i.d. zero-mean, unity-variance random variables (RVs), it can be shown by the law of large numbers that $\lim_{n \to \infty} \frac{1}{n} p^t q \to 1$ and $\lim_{n \to \infty} \frac{1}{n} p^t q \to 0$ where $\frac{1}{n} \to$ denotes the almost sure convergence.

For perfect CSI using MRC, $f_{n_g,k} = h_{n_g,k}^H$ [3], and the SINR of user $n_g$ on sub-carrier $k$ is $\gamma_{n_g,k} = P_{n_g,k}\|h_{n_g,k}\|^2/(\|I_{n_g,k}\|^2 + \|h_{n_g,k}\|^2)$. According to the law of large numbers for large numbers of large $M_n_g,k$, $\|h_{n_g,k}\|^2 \to M_n_g,k d_n_g$ and $\|I_{n_g,k}\|^2 \to 0$. Thus, $R^\text{Perf}_{n_g,k} = \log_2 (1 + P_{n_g,k} d_n_g M_n_g,k)$ as shown in (3).

For imperfect CSI in (2), $\delta_{n_g,k}(P) = \sum_{g \in G} \sum_{n_g \neq n_g} P_{n_g,k} h_{n_g,k} e_{n_g,k} h_{n_g,k}^H x_{m_g,k} + z_{n_g,k}^\text{Imperf}_{n_g,k}$

where $e_{n_g,k} = h_{n_g,k} - \hat{h}_{n_g,k}$ whose elements are RVs with zero mean and variance $\hat{P}_{n_g,k} d_n_g + \sigma_{n_g,k}^2 = \tau P_{n_g,k}$ [3].

With MMSE-based channel estimation, the elements of $\hat{h}_{n_g,k}$ are i.i.d. RVs with zero mean and variance $\hat{P}_{n_g,k} d_n_g + \sigma_{n_g,k}^2 = \tau P_{n_g,k}$ [3].

With MRC precoder, $\gamma_{n_g,k}^\text{Imperf} = \frac{P_{n_g,k}}{\hat{P}_{n_g,k} d_n_g + \sigma_{n_g,k}^2} \frac{\|h_{n_g,k}\|^2}{\|\hat{h}_{n_g,k}\|^2}$. Since $e_{n_g,k}$ and $h_{n_g,k}$ are independent of $h_{n_g,k}$ and $\hat{h}_{n_g,k}$, the first term of $\delta_{n_g,k}(P)$ is zero and second term is equal to $\sqrt{\frac{P_{n_g,k}}{\hat{P}_{n_g,k} d_n_g + \sigma_{n_g,k}^2}} \frac{\|h_{n_g,k}\|^2}{\|\hat{h}_{n_g,k}\|^2} \frac{\|\hat{h}_{n_g,k}\|^2}{\|h_{n_g,k}\|^2}$

Thus, the SINR for the imperfect CSI case is $\gamma_{n_g,k} = P_{n_g,k} \|\hat{h}_{n_g,k}\|^4/(\|P_{n_g,k} \|e_{n_g,k}^H h_{n_g,k}\|^2 + \|\hat{h}_{n_g,k}\|^2)^2$

By some mathematical manipulations, $\gamma_{n_g,k}^\text{Imperf} = \frac{M_n_g,k d_n_g + \tau P_{n_g,k}^2}{P_{n_g,k} d_n_g + \tau P_{n_g,k}^2 + \sigma_{n_g,k}^2}$. Now, by substituting $P_{n_g,k}^\text{Pilot} = \tau P_{n_g,k}$ and considering $P_{n_g,k} = \rho_{n_g,k}/\sqrt{M_n_g,k}$ [3], we have $\gamma_{n_g,k}^\text{Imperf} = \frac{M_n_g,k d_n_g + \tau P_{n_g,k}^2}{\rho_{n_g,k} d_{n_g} + \tau P_{n_g,k}^2 + \sigma_{n_g,k}^2}$. When $M_n_g,k \to \infty$, we have $\gamma_{n_g,k}^\text{Imperf} = \tau \rho_{n_g,k}^2 d_{n_g}^2 = \tau P_{n_g,k}^2 d_{n_g}^2 M_n_g,k$. Thus, as shown in (3), $R_{n_g,k} = \frac{1}{T} \log_2 (1 + \tau P_{n_g,k}^2 d_{n_g}^2 M_n_g,k)$ where $(T - \tau)/T$ is the fraction of transmission frame for sending the actual data [3].

REFERENCES