A Utility-Based Resource Allocation in Virtualized Cloud Radio Access Network

Linna Chen(✉), Chunjing Hu, Yong Li, and Wenbo Wang

Wireless Signal Processing and Network Laboratory,
Key Laboratory of Universal Wireless Communications, Ministry of Education,
Beijing University of Posts and Telecommunications, Beijing 100876, China
chenlinna@bupt.edu.cn

Abstract. Network slicing is an emerging paradigm for 5G networks. Network slices are considered as different and independent virtualized end-to-end networks on a common physical infrastructure. Wireless resource virtualization is the key enabler to achieve high resource efficiency and meanwhile to isolate network slices from one another. In this paper, we propose a slice-specific utility-based resource allocation scheme in cloud radio access networks, where two sets of slices with different requirements are supported simultaneously. Every slice can determine its preference factor in utility function considering the trade-off between bandwidth gain and energy consumption. The objective is to maximize the sum utility of all slices taking the trade-off of all slices into account, which can be formulated as a mixed binary integer nonlinear programming problem. The Lagrange dual method is applied to solve the joint optimization problem. Finally, The performance of the proposed scheme is evaluated and the results show that the proposed scheme can meet different customized requirements of all slices, and enhance system performance when compared with other methods.

Keywords: Network slicing · Virtualization · Utility · Trade-off

1 Introduction

With rapid increase of traffic volumes and demand for a variety of services, the traditional one-size-fits-all network architecture is hard to adapt to the requirements of emerging use cases and changing subscriber demands. Network slicing is an emerging paradigm for 5G networks, which will enable operators to provide networks on an as-a-service basis and meet the wide range of use cases that the 2020 timeframe will demand. Each service provider (SP) can require its own logical network slice to support specific subscriber types and varying application usages on a shared physical infrastructure. In order to handle multiple slices in a robust way, radio access networks (RANs) shall provide radio resource management to efficiently multiplex traffics from multiple users in different network slice instances onto the available shared radio resources.

K. Long et al. (Eds.): 5GWN 2017, LNICST 211, pp. 288–299, 2018.
https://doi.org/10.1007/978-3-319-72823-0_28

Wireless network virtualization is an effective means by which an infrastructure provider can partition the wireless and physical resources among SPs so that they can serve their subscribers, which facilitates new business models. Due to resource and infrastructure sharing, wireless network virtualization can reduce capital expenditure and operational expenditure, which leads to potential energy and capital savings [1]. Moreover, wireless network virtualization contributes to better resource utilization due to less unused resource.

The problem of wireless virtualization in LTE is addressed in many papers. The authors in [2] proposed a slicing scheme to allocate physical resources of LTE system to different SPs, in which fairness requirements of different SPs were considered. In [3], a centralized heuristic to allocate radio resource blocks in multi-cell LTE networks was proposed, which aimed to maximize the sum rate of the network. In [4,5], a resource allocation scheme was proposed by introducing two types of slices, including rate-based slices and resource-based slices. In [6], the authors introduced an idea of wireless virtualization into full-duplex relaying networks and proposed a virtual resource management architecture for virtualized networks. However, the wireless virtualization in cloud RAN (CRAN) has not been discussed adequately.

In this paper, we consider the joint subcarrier and power allocation problem in a CRAN downlink system, where two sets of slices with different requirements are supported simultaneously. Our main contribution is to introduce a slice-specific utility-based resource allocation scheme. The scheme aims at maximizing the sum utility of all slices under the service level agreement (SLA) and QoS requirement constrains, in which the specific preference requirements of different slices are considered. Every slice can determine a customized weighing factor in its utility function considering the trade-off between spectral efficiency and energy efficiency. The proposed allocation algorithm will take the preference requirements of all slices into account when assigning the subcarriers and power between different slices.

The rest of this paper is organized as follows. First, we present a description of the system model and the problem formulation in Sect. 2. The approach to solve the optimal problem is presented in Sect. 3. The simulation results and their discussions are given in Sect. 4. Finally, we conclude this paper in Sect. 5.

2 System Model and Problem Formulation

We consider the downlink of the CRAN architecture, where the coverage of a certain geographical area is provided by a cluster of RRHs, as illustrated in Fig. 1.

2.1 OFDMA-Based Wireless Transmission

We assume that OFDMA is used for downlink transmission. The total channel bandwidth is B Hz and is divided into N orthogonal subcarriers, thus the bandwidth of each subcarrier is B/N. The system consists of G slices, where each

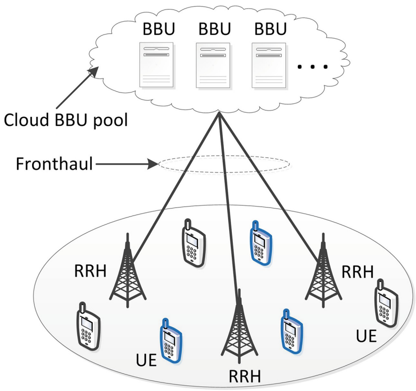

Fig. 1. System model.

slice g provides its service to K_g users, and each user is cooperatively served by its serving cluster of R RRHs.

Each RRH has a total transmit power of $P_{j,max}$ and let $p_{i,j,n}$ denote the power allocated to user i on subcarrier n at RRH j. The bandwidth of each subcarrier is assumed to be small compared with the coherent bandwidth of the wireless channel. Therefore, $h_{i,j,n}$ is the channel gain of the wireless link between RRH j and user i on subcarrier n which can be considered flat. Then the baseband complex symbol $y_{i,n}$ received by user i at subcarrier n can be expressed as

$$y_{i,n} = \sum_{j=1}^{R} h_{i,j,n} w_{i,j,n} \sqrt{p_{i,j,n}} s_{i,n} + z_{i,n}, \tag{1}$$

where $s_{i,n} \sim \mathcal{CN}(0,1)$ is the signal transmitted to UE i at subcarrier n, and $w_{i,j,n} = \frac{h^*_{i,j,n}}{|h_{i,j,n}|}$ denotes the complex precoding symbol, by which the phases of transmission signals from different RRHs could be rotated into the same direction as the phase of the initial transmission signal. Also $z_{i,n} \sim \mathcal{CN}(0, \sigma^2)$ denotes the received additive white Gaussian noise (AWGN), where σ^2 is the noise variance. Let $R_{i,n}$ denote the transmission rate of user i on subcarrier n, which can be calculated as

$$\begin{aligned} R_{i,n} &= \log_2\left(1 + \frac{|\sum_{j=1}^{R} h_{i,j,n} w_{i,j,n} \sqrt{p_{i,j,n}}|^2}{\sigma^2}\right) \\ &= \log_2\left(1 + \frac{|\sum_{j=1}^{R} (|h_{i,j,n}| \sqrt{p_{i,j,n}})|^2}{\sigma^2}\right). \end{aligned} \tag{2}$$

2.2 Problem Formulation

The system provides services for two specific sets of slices: (1) One set with specific QoS requirements, the slice g in this set requires a minimum reserved rate $R_{g,min}$, $g = 1, 2, \dots G1$; (2) The other set without QoS requirements, the traffic of slice g can be delivered in a best-effort manner, $g = G1 + 1, \dots G$. Meanwhile, each slice must be assigned with a minimum amount of resources to guarantee the SLA, which can ensure isolation between slices to a certain extent. For example, when a slice is overloaded, the other slices could still obtain its certain amount of resources.

In this paper, a slice-specific utility-based resource allocation is proposed. Utility function varies across slices with different requirements considering the trade-off between the gain on throughput and the cost on power consumption, where ε_g denotes the preference coefficient. The preference coefficient reflects the specific need of the individual slice toward the above two types of aspects in the process of resource allocation. For example, with a higher ε_g, the slice needs to pay a higher cost for the allocated power, which means that the slice prefers to minimize the power consumption. Otherwise, the slice may be less concerned on power consumption but more concerned on the throughput gain. The scheduling policy would consider the preferences of all slices to allocate the resources between different slices properly. Let $x_{i,n} \in \{0, 1\}$ denote the subcarrier allocation binary variable, where $x_{i,n} = 1$ indicates that subcarrier n is assigned to user i, and otherwise $x_{i,n} = 0$. Thus the utility function of slice g is defined as follows

$$U_g = \sum_{i=1}^{K_g} \sum_{n=1}^{N} x_{i,n} R_{i,n} - \varepsilon_g \sum_{j=1}^{R} \sum_{i=1}^{K_g} \sum_{n=1}^{N} x_{i,n} p_{i,j,n}. \tag{3}$$

Therefore the problem we considered is to optimize the allocation of subcarriers and power jointly so as to maximize the total utility of all slices subject to the physical limitations and SLA of slices, which can be formulated as

$$\begin{aligned}
&\max_{\mathbf{x},\mathbf{p}} \sum_{g=1}^{G} U_g \\
&\text{s.t.} \\
&\text{C1}: \sum_{g=1}^{G} \sum_{i=1}^{K_g} x_{i,n} \le 1, \forall n, \\
&\text{C2}: \sum_{g=1}^{G} \sum_{i=1}^{K_g} \sum_{n=1}^{N} x_{i,n} p_{i,j,n} \le P_{j,max}, \forall j, \\
&\text{C3}: \sum_{i=1}^{K_g} \sum_{n=1}^{N} x_{i,n} R_{i,n} \ge R_{g,min}, \forall g = 1, 2 \dots G1, \\
&\text{C4}: \sum_{i=1}^{K_g} \sum_{n=1}^{N} x_{i,n} \ge N \rho_{g,min}, \forall g = 1, 2 \dots G,
\end{aligned} \tag{4}$$

where C1 represents the exclusive orthogonal constraint which ensures that each subcarrier is allowed to be assigned to one user at most, and C2 is the individual power constraint of each RRH, and C3 represents the minimum required rate of each slice g in the set with QoS requirements, and C4 ensures the isolation between slices such that each slice can access at least a certain number of subcarriers to guarantee its SLA, where $\rho_{g,min}$ denotes the contracted minimum portion of resources assigned to slice g, $\rho_{g,min} \in [0,1]$,$\forall g$ and $\sum_{g=1}^{G} \rho_{g,min} \leq 1$.

3 Problem Solution and Allocation Algorithm

3.1 Problem Solution

The objective function in (4) with its constraints can be classified as a mixed binary integer nonlinear programming problem, with the decision variables $x_{i,n}$ and $p_{i,j,n}$ being binary variable and continuous variable, respectively. The complexity of solving this problem is high, so we will propose an approach to make the problem tractable using the dual decomposition method, similar to the techniques used in [7–9]. Consequently, the Lagrange function of the above optimization problem is given by

$$\begin{aligned} L(\mathbf{x}, \mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{u}) = & \sum_{g=1}^{G} \left[\sum_{i=1}^{K_g} \sum_{n=1}^{N} x_{i,n} R_{i,n} - \varepsilon_g \sum_{j=1}^{R} \sum_{i=1}^{K_g} \sum_{n=1}^{N} x_{i,n} p_{i,j,n} \right] \\ & + \sum_{n=1}^{N} \lambda_n (1 - \sum_{g=1}^{G} \sum_{i=1}^{K_g} x_{i,n}) \\ & + \sum_{j=1}^{R} \mu_j (P_{j,max} - \sum_{g=1}^{G} \sum_{i=1}^{K_g} \sum_{n=1}^{N} x_{i,n} p_{i,j,n}) \\ & + \sum_{g=1}^{G} \nu_g (\sum_{i=1}^{K_g} \sum_{n=1}^{N} x_{i,n} R_{i,n} - R_{g,min}) \\ & + \sum_{g=1}^{G} u_g (\sum_{i=1}^{K_g} \sum_{n=1}^{N} x_{i,n} - N \rho_{g,min}), \end{aligned} \tag{5}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_R)$, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_G)$, $\mathbf{u} = (u_1, \ldots, u_G)$ are the non-negative Lagrange multipliers for the constraints C1 - C4, respectively, and $\nu_g = \nu_g$ when $1 \leq g \leq G1$, or $\nu_g = 0$ when $G1 < g \leq G$. Herein, we can give the dual objective function as

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{u}) = \max_{\mathbf{x},\mathbf{p}} L(\mathbf{x}, \mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{u}). \tag{6}$$

The dual optimization problem is then formulated as

$$\begin{aligned} & \min_{\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{u}} g(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{u}) \\ & \text{s.t. } \boldsymbol{\lambda} \succeq 0, \boldsymbol{\mu} \succeq 0, \boldsymbol{\nu} \succeq 0, \mathbf{u} \succeq 0. \end{aligned} \tag{7}$$

To solve the dual problem, we can re-express the dual objective function as

$$
\begin{aligned}
g(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{u}) = & \sum_{n=1}^{N} g_n(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{u}) + \sum_{n=1}^{N} \lambda_n + \sum_{j=1}^{R} \mu_g P_{j,max} \\
& - \sum_{g=1}^{G} \nu_g R_{g,min} - \sum_{g=1}^{G} u_g N \rho_{g,min},
\end{aligned} \tag{8}
$$

where

$$
g_n(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{u}) = \max_{\mathbf{x},\mathbf{p}} \left[\sum_{g=1}^{G} \sum_{i=1}^{K_g} x_{i,n} g_{i,n}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{u}) \right], \tag{9}
$$

$$
g_{i,n}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{u}) = R_{i,n} - \varepsilon_g \sum_{j=1}^{R} p_{i,j,n} - \lambda_n - \sum_{j=1}^{R} \mu_j p_{i,j,n} + \nu_g R_{i,n} + u_g. \tag{10}
$$

We have decomposed the dual function into the above N independent optimization sub-problems given by (9). By evaluating the Hessian matrix of $R_{i,n}$ at $p_{i,j,n}$ in (2), we can prove that the function in (2) is concave. Thus, the function in (10) is concave as any positive linear combination of concave functions is concave. Suppose subcarrier n is assigned to user i, the optimal $p^*_{i,j,n}$ that maximizes the object function in (10) for fixed Lagrange multipliers can be easily obtained. Then by comparing all possible user assignments of this subcarrier, we select user i for which $g_{i,n}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{u})$ is maximized, and allocate subcarrier n to that user. The allocation of all N subcarriers can be obtained in the same way.

Once the optimal $\mathbf{x}^*$, $\mathbf{p}^*$ are obtained, we can use them to solve the dual problem in (7) to find the optimal values of dual variables. Note that the Lagrange function $L(\mathbf{x}, \mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{u})$ is linear in $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$, $\boldsymbol{\nu}$, $\mathbf{u}$ for fixed $\mathbf{x}^*$ and $\mathbf{p}^*$, and $g(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{u})$ is the maximum of these linear functions, so the dual problem in (7) is convex. With the help of the sub-gradient method, we can find the optimal values of dual variables, which can be given by

$$
\mu_j^{t+1} = \left[\mu_j^t - \delta_j (P_{j,max} - \sum_{g=1}^{G} \sum_{i=1}^{K_g} \sum_{n=1}^{N} x^*_{i,n} p^*_{i,j,n}) \right]^+, \tag{11}
$$

$$
\nu_g^{t+1} = \left[\nu_g^t - \xi_g (\sum_{i=1}^{K_g} \sum_{n=1}^{N} x^*_{i,n} R^*_{i,n} - R_{g,min}) \right]^+, \tag{12}
$$

$$
u_g^{t+1} = \left[u_g^t - \zeta_g (\sum_{i=1}^{K_g} \sum_{n=1}^{N} x^*_{i,n} - N \rho_{g,min}) \right]^+, \tag{13}
$$

where δ_j, ξ_g and ζ_g are the appropriate small step sizes to guarantee the convergence of the sub-gradient method. The iteration process will be stopped until certain criterion is fulfilled.

3.2 An Iterative Algorithm

The optimization problem is typically required to be decomposed into N sub-problems to reduce its complexity. After obtaining the optimal dual variables that minimize the dual function, it remains to find the optimal primal solutions $\mathbf{x}^*$, $\mathbf{p}^*$ that maximize the Lagrangian function and satisfy all constraints in the original problem (4). The final optimal solution would be achieved in such an iterative manner. Our proposed algorithm is proposed as follows:

1. Initialize the optimal variables $\mathbf{x}^0$, $\mathbf{p}^0$ and the dual variables $\boldsymbol{\lambda}^0$, $\boldsymbol{\mu}^0$, $\boldsymbol{\nu}^0$, $\mathbf{u}^0$.
2. For each iteration t, solve the N sub-problems in the following steps:
   (a) For the given dual variables, compute the optimal $p^*_{i,j,n}$ that maximizes the object function in (10).
   (b) Update the optimal $g^*_{i,n}$ for every user i.
   (c) Select $k = \arg\max g^*_{i,n}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{u})$, and set $x_{k,n} = 1$, otherwise $x_{i,n} = 0$, for $i \neq k$.
   (d) Once the assignment problems are solved for all N subcarriers, the optimal variables $\mathbf{x}^*$, $\mathbf{p}^*$ can be obtained.
3. Update the $\boldsymbol{\lambda}^{t+1}$, $\boldsymbol{\mu}^{t+1}$, $\boldsymbol{\nu}^{t+1}$, $\mathbf{u}^{t+1}$ using the obtained $\mathbf{x}^*$ and $\mathbf{p}^*$ and let $t = t + 1$.
4. Continue to the next iteration in step (2) until convergence or the maximum iteration number t_{max} is reached.

3.3 Complexity of the Algorithm

The optimal solution to (10) can be obtained by using global searching, assuming that each $p_{i,j,n}$ takes discrete values and requires $\mathcal{O}(X)$ computational complexity. Thus, the optimal power allocation solution requires $\mathcal{O}(X^R)$ computational complexity. For the given dual variables, the complexity of updating $\mathbf{x}$ in each iteration is $\mathcal{O}(NGK_gX^R)$. Let L be the number of iterations needed to converge in the sub-gradient method. Therefore the total computation complexity of the proposed algorithm is $\mathcal{O}(LNGK_gX^R)$.

4 Numerical Results

4.1 System Setup

In this section, the performance evaluation for the proposed allocation algorithm is presented. There are 3 RRHs considered to cover a certain geographical area. Without loss of generality, the channels of different subcarriers are assumed to be independent of one another, which are taken from i.i.d. complex Gaussian random variables with zero mean and unit variance, and the noise variance is given as $\sigma^2 = 0.1$. The number of subcarriers is taken as 32. We assume that the system consists of two slices, each of which serves 10 subscribers. One slice (Slice 1) has QoS requirement that needs a minimum rate of 120 bit/Hz, and the other one (Slice 2) has no QoS requirement. Meanwhile, the contracted minimum resources of each slice is given as $\rho_{g,min} = 0.25$. For comparison, we use two baseline schemes. The first scheme is equal power allocation (EPA) algorithm where the total power is shared equally among subcarriers, and the second scheme is static sharing (SS) scheme, which statically assigns fixed subcarriers to each slice.

4.2 Results and Discussion

In Fig. 2, we present the total achievable system throughput of the proposed scheme, EPA scheme and SS scheme, respectively. In this case, the preference coefficient ε_g is set to zero for each slice, which means that both slices would like to maximize the bandwidth gain no matter how much power is consumed. We can see that, the maximum achievable throughput of all three schemes increases monotonically with P_{max}, and our proposed scheme outperforms all the other schemes. SS scheme offers the lowest throughput since users of each slice can only access their dedicated subcarriers, and have no chance to use underutilized resources that belong to other slice. Figure 3 depicts the impact of different QoS constraints of Slice 1 on sum utility of the system. In this case, the preference factors are set as $\varepsilon_1 = 0.6$ for Slice 1, and $\varepsilon_2 = 0.2$ for Slice 2. The results show that the sum utility decreases as the rate constraint of Slice 1 increases. This is because, by increasing R_{min}, more transmit power should be consumed to satisfy the QoS constraints. And the allocated subcarriers for users with best channel conditions in Slice 2 decreases when the Qos requirement of Slice 1 increases which leads to inefficient use of the resources hence lowering the sum utility. Thus the system can hold the strict rate constraint at the cost of sum utility. In addition, at the same R_{min}, the utility of the proposed scheme is larger than that of EPA scheme since our proposed scheme is a joint subcarrier and power allocation algorithm but EPA scheme is not.

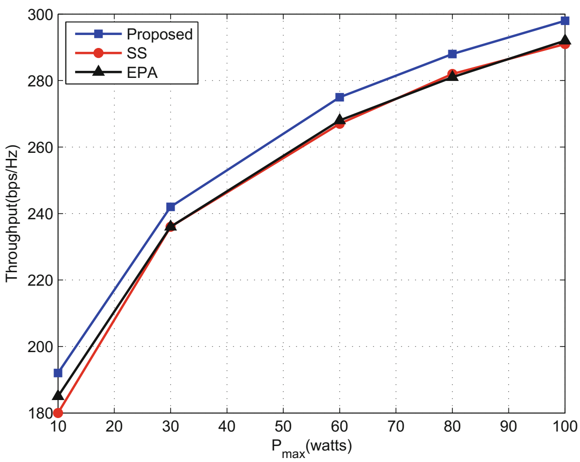

Fig. 2. Throughput of the system versus maximum transmit power P_{max}.

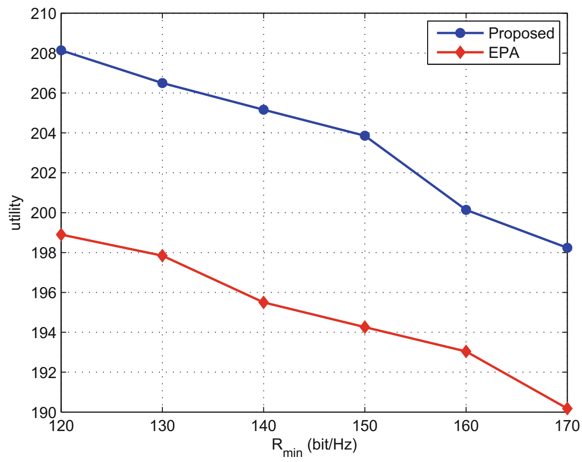

Fig. 3. The total utility of the system versus QoS constrain of Slice 1 R_{min}.

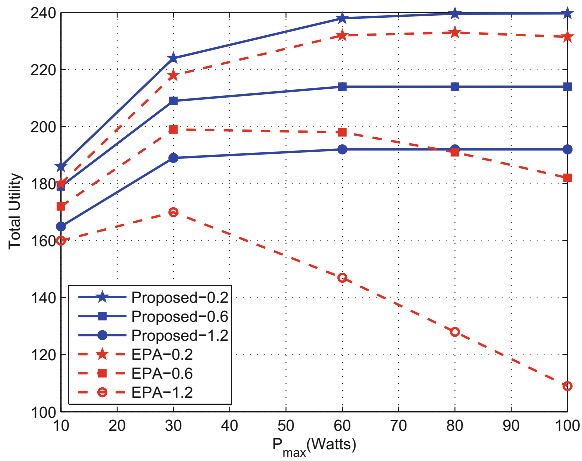

Fig. 4. Total utility versus P_{max} with different factor of Slice 1.

Figure 4 illustrates the impact of the maximum transmit power of each RRH on sum utility of the system. The preference coefficient of Slice 2 is given by $\varepsilon_2 = 0.2$, while Slice 1 has three coefficient sets, $\varepsilon_1 = 0.2$, $\varepsilon_1 = 0.6$ and $\varepsilon_1 = 1.2$ for comparison. It can be noticed that the proposed scheme has better performance than the EPA scheme in all three cases, and the system utility of a higher ε_1 is worse than that of a lower ε_1, since a higher ε_1 means that Slice 1 must pay much more for the power allocated in order to achieve the same throughput constraint. Further, we can see that, When the transmit power is relatively low, the maximum achievable utility increases monotonically with the increasing power in all schemes. The optimal utility can be achieved with a high power constraint P_{max}. In this case, the achievable utility of the proposed scheme no longer increases with the constraint P_{max} as no more power is consumed. However, the EPA scheme continues to allocate more power in the high P_{max} region, resulting in the system utility dropping dramatically, especially when the preference coefficient is higher. This is because, the system utility has a diminishing return with respect to the increase of the transmit power, higher power consumption counteracts the throughput gain.

To study the impact of different preference coefficients to the slice performance, we take "bit/Hz/Joule" as the metric for EE [10], and EE is defined as $\frac{R_g}{P_g}$ where R_g denotes the achieved throughput of the slice g, P_g denotes its consumed power. In this case, the preference factor of Slice 2 is fixed as $\varepsilon_2 = 0.2$, while the factor of Slice 1 has three options: $\varepsilon_1 = 0.2$, $\varepsilon_1 = 0.6$ or $\varepsilon_1 = 1.2$. From Fig. 5, we can observe that the higher ε_1 leads to higher slice-EE of Slice 1. This is because the higher ε_1 means that Slice 1 prefers to achieve the similar throughput in a more energy-efficient way, which can also easily explain why the EE of Slice 1 is much higher than that of Slice 2. In addition, it can be observed in Fig. 6 that the EE of Slice 2 is lower when the preference factor of Slice 1 is higher. Because Slice 1 is more concerned on power consumption than Slice 2 in such case, and the scheduling policy will consider the preference of both slices when allocating resources between slices in order to maximize the utility of whole system. Therefore the resource allocation is energy-efficient to Slice 1, while it may be opposite from the perspective of Slice 2. However, the two slices can only achieve very similar slice-EE in EPA scheme regardless of the preference coefficients they chose in their utility function, since the power allocation is unchanged in this scheme. Finally, Our proposed scheme can satisfy the different requirements of slices according to their trade-off between throughput gain and power consumption, which can achieve better system performance.

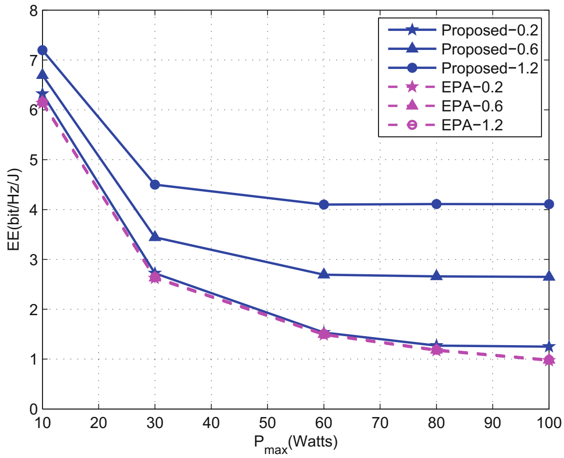

Fig. 5. EE of Slice 1 versus P_{max} for different factor of Slice 1.

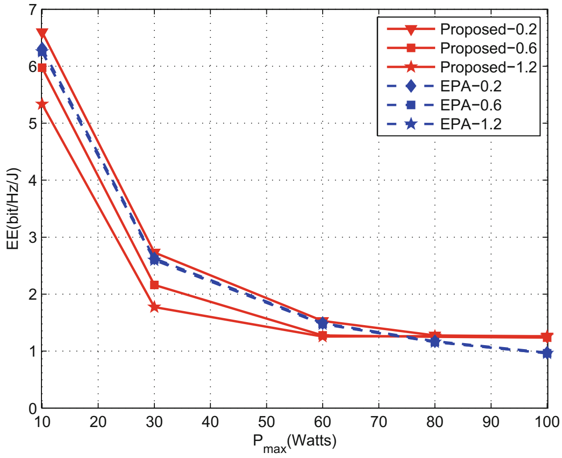

Fig. 6. EE of Slice 2 versus P_{max} for different factor of Slice 1.

5 Conclusions

In this paper, we studied slice-specific utility-based resource allocation in CRANs. In the proposed resource allocation algorithm, each slice has a customized utility function taking their specific trade-off between spectral efficiency and energy efficiency into account. The objective function is to maximize the sum utility of all slices through joint subcarriers and power allocation under different QoS requirement constraints of slices. Numerical results demonstrated that the proposed algorithm significantly outperforms other candidates, which can satisfy the customized preference of slices in terms of throughput gain and power cost.

Acknowledgement. The authors would like to thank DOCOMO Beijing Communications Laboratories Co., Ltd. for their support in this work. This work was supported in part by the National High Technology Research and Development Program (863 Program) of China under Grant No. 2014AA01A707.

References

1. Kalil, M., Shami, A., Ye, Y.: Wireless resources virtualization in LTE systems. In: Proceedings of IEEE INFOCOM Workshops 2014, Toronto, Ontario, Canada, pp. 363–368, April 2014
2. Kamel, M.I., Le, L.B., Girard, A.: LTE wireless network virtualization: dynamic slicing via flexible scheduling. In: Proceedings of IEEE VTC Fall 2014, Vancouver, British Columbia, Canada, pp. 1–5, September 2014
3. Kamel, M.I., Le, L.B., Girard, A.: LTE multi-cell dynamic resource allocation for wireless network virtualization. In: Proceedings of IEEE WCNC 2015, New Orleans, Louisiana, USA, pp. 966–971, March 2015

4. Kokku, R., Mahindra, R., Zhang, H., Rangarajan, S.: NVS: a substrate for virtualizing wireless resources in cellular networks. IEEE/ACM Trans. Netw. **20**(5), 1333–1346 (2012)
5. Parsaeefard, S., Jumba, V., Derakhshani, M., Le-Ngoc, T.: Joint resource provisioning and admission control in wireless virtualized networks. In: Proceedings of IEEE WCNC 2015, New Orleans, Louisiana, USA, pp. 2020–2025, March 2015
6. Liu, G., Yu, F.R., Ji, H., Leung, V.C.M.: Distributed resource allocation in virtualized full-duplex relaying networks. IEEE Trans. Veh. Technol. **65**(10), 8444–8460 (2016)
7. Palomar, D.P., Chiang, M.: A tutorial on decomposition methods for network utility maximization. IEEE J. Sel. Areas Commun. **24**(8), 1439–1451 (2006)
8. Yu, W., Lui, R.: Dual methods for nonconvex spectrum optimization of multicarrier systems. IEEE Trans. Commun. **54**(7), 1310–1322 (2006)
9. Palomar, D.P., Chiang, M.: Alternative decompositions for distributed maximization of network utility: framework and applications. In: Proceedings of IEEE INFOCOM 2006, Barcelona, Spain, pp. 1–13, April 2006
10. Zou, J., Xi, Q., Zhang, Q., He, C., Jiang, L., Ding, J.: QoS-aware energy-efficient radio resource allocation in heterogeneous wireless networks. In: Proceedings of IEEE ICCW 2015, London, UK, pp. 2781–2786, June 2015