Dynamic resource provisioning with stable queue control for wireless virtualized networks

Dynamic Resource Provisioning with Stable Queue Control for Wireless Virtualized Networks

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; saeideh.parsaeifard@mcgill.ca; mahsa.derakhshani@utoronto.ca; tho.le-ngoc@mcgill.ca

Abstract—**This paper investigates the dynamic resource provisioning with queue stability in wireless virtualized networks (WVN). Aiming to maximize the total average rate of WVN over a transmission frame, a dynamic resource provisioning policy is proposed, while a minimum average required rate of each slice and a stable-queue constraint of WVN are preserved. Based on Lyapunov drift-plus-penalty algorithm and variable transformation techniques, an iterative algorithm is proposed for joint power and sub-carrier allocation. Performance of the proposed algorithm is evaluated by simulations performed investigate the effects of various system parameters on the average rate of WVN and queue stability.**

Index Terms—**Dynamic resource provisioning, Lyapunov drift-plus-penalty algorithm, network queue stability, virtualized wireless networks.**

I. INTRODUCTION

The recent trend of service-oriented demands of mobile users calls for more efficient and modular wireless networks. The concept of wireless virtualization is a solution to provide flexibility and service customization in the existing networks via sharing the limited wireless resources (e.g., power and sub-carriers) among various service providers (also referred to as slices). Generally, each slice has its own set of users and supports specific quality-of-service (QoS) requirements for offered services.

Considering dynamic features of wireless channels (e.g., fading and shadowing) and limited resources in WVN, an effective dynamic resource provisioning is required to satisfy the physical-layer QoS requirements of each slice, e.g., minimum required rate of each slice, subject to WVN limitations e.g., transmit power at base stations (BSs). Recently, this problem has received a lot of attention. For instance, in [1], a resource management scheme is studied by introducing rate-based and resource-based slices requiring minimum rate and physical resources, respectively. In [2], an auction game is applied for the interaction of slices, network operator, and users. To guarantee the minimum rate of slices under fading of wireless channels, an admission control policy is proposed in [3] where the minimum rate of each slice is adjusted based on its users' channel state information (CSI). To extend the feasibility condition of WVN for supporting diverse QoS, [4] borrowed the concept of massive MIMO. In [5], an elastic resource allocation problem is combined with time and space for OFDMA system.

However, in addition to the physical-layer (PHY) parameters considered by the above-mentioned works, it is very important to consider random traffic arrivals and the limited queue backlogs of users aiming to reach reliable service support for each user. To address this issue, we propose the dynamic resource provisioning algorithm based on both PHY parameters (i.e., CSI) and the medium access layer (MAC) parameters (i.e., queue state information (QSI)). The main aim of this optimization problem is to maximize the total rate of WVN subject to average queue stability condition of each user and average minimum required rate of each slice over one complete transmission frame including a number of time slots.

The proposed optimization problem of this paper is challenging to solve because of its cross-layer nature and both PHY and MAC constraints. To tackle this, we apply the Lyapunov drift-plus-penalty algorithm [6] to transform the average rate dynamic problem over one transmission frame into one time slot optimal problem. Lyapunov drift approach and drift-plus-penalty algorithm have been widely applied for the resource allocation problem of conventional wireless networks by considering limited queue length or maximum delay requirement, e.g., [7]–[14]. However, to the best of our knowledge, there is no other related work that apply this framework to solve the dynamic resource allocation problem in WVN.

In the down-link orthogonal frequency division multiple access (OFDMA) based transmission scenario, the optimization problem of each time slot includes both sub-carrier assignment and power allocation variables. Therefore, proposing the efficient algorithm to get rid of high computational complexity is of high importance. To this end, we first apply the technique of variable relaxations and transformations [15] to transform non-convex problem into a related convex one. Afterward, we propose a two-level iterative algorithm to solve the queue-aware dynamic resource provisioning: 1) Outer loop, where queue parameters are updated based on the achieved rate of slices. 2) Inner loop, where power and sub-carriers are derived by applying KKT conditions.

The rest of this paper is organized as follows. Section II introduces the system model and problem formulations followed by Section III, which provides detailed solution for the problem via Lyapunov drift-plus-penalty approach and eventually proposes the iterative algorithm and its performance analysis. Section IV presents the simulation results and Section V concludes the paper.

II. Network Model and Dynamic Resource Provisioning Problem

We consider a slotted down-link OFDMA-based WVN which provides the coverage to a specific region by one base station (BS). In this area, WVN serves a set of $\mathcal{G} = \{1, ..., G\}$ slices. Each slice $g \in \mathcal{G}$ serves a group of users denoted by $\mathcal{N}_g = \{1, \cdots, N_g\}$. The total number of users is $N = \sum_{g \in \mathcal{G}} N_g$.

The network operates in slotted transmission time $\mathcal{T} = \{1, \cdots, t, \cdots, T\}$. The total bandwidth of B Hz is shared between users of all slices through OFDMA scheme, which is divided into a set of $\mathcal{K} = \{1, \cdots, K\}$ sub-carriers. The bandwidth of each sub-carrier, $B_c = B/K$, is assumed to be small compared to the coherent bandwidth of the wireless channel. Based on this assumption, consider $h_{n_g,k}^t$ as the channel power gain of user n_g on sub-carrier $k \in \mathcal{K}$ in time slot $t \in \mathcal{T}$. Also, consider $p_{n_g,k}^t$ as the allocated power to user n_g on sub-carrier k in time slot t and $w_{n_g,k}^t = \{0,1\}$ as the sub-carrier assignment variable where $w_{n_g,k}^t = 1$ if the sub-carrier k is allocated to user n_g at time slot t, otherwise $w_{n_g,k}^t = 0$. In this setup, the total transmit power at the BS is limited to the maximum power $P_{\max}$ in each time slot, i.e.,

$$\text{C1:} \quad \sum_{g \in \mathcal{G}} \sum_{n_g \in \mathcal{N}_g} \sum_{k \in \mathcal{K}} w_{n_g,k}^t p_{n_g,k}^t \leq P_{\max}.$$

Furthermore, from OFDMA limitation, each sub-carrier is assigned exclusively to one user where

$$\text{C2}: \sum_{g \in \mathcal{G}} \sum_{n_g \in \mathcal{N}_g} w_{n_g,k}^t \leq 1 \text{ and } w_{n_g,k}^t \in \{0,1\}, \forall k \in \mathcal{K}.$$

Without loss of generality, let σ be the noise power in each sub-carrier. In this scenario, the rate of user $n_g \in \mathcal{N}_g$ at time slot t is

$$R_{n_g}^t(\mathbf{P}_{n_g}, \mathbf{w}_{n_g}) = \sum_{k \in \mathcal{K}} w_{n_g,k}^t \log_2 \left(1 + \frac{p_{n_g,k}^t h_{n_g,k}^t}{\sigma} \right),$$

where $\mathbf{P}_{n_g} = [p_{n_g,k}^t]_{k=1}^K$, $\mathbf{w}_{n_g} = [w_{n_g,k}^t]_{k=1}^K$ and

$$\bar{R}_{n_g} = \lim_{T \to \infty} \frac{1}{T} \sum_{t \in T} R_{n_g}^t(\mathbf{P}_{n_g}, \mathbf{w}_{n_g})$$

is the average rate of user n_g over a transmission frame. Now, the minimum required average rate of slice $g \in \mathcal{G}$ over one transmission frame, $\bar{R}_g^{\text{rsv}}$ is presented as

$$\text{C3:} \quad \sum_{n_g \in \mathcal{N}_g} \bar{R}_{n_g} \geq \bar{R}_g^{\text{rsv}}, \ \forall g \in \mathcal{G}.$$

Let $\alpha_{n_g}^t$ and $Q_{n_g}^t$ respectively be the i.i.d. packet arrival process and queue length for user n_g in time slot t. Furthermore, let l be the size of a packet in bits per Hz per slot. Therefore, the queue dynamics for user n_g can be mathematically represented as

$$Q_{n_g}^{t+1} = \max [Q_{n_g}^t - R_{n_g}^t(\mathbf{P}_{n_g}, \mathbf{w}_{n_g}) T_s, 0] + \alpha_{n_g}^t, \quad (1)$$

where T_s is the duration of each time slot $t \in \mathcal{T}$. In this

setup, network has a queue stability condition if and only if the queue of each user is stable. The queue of each user is stable when *"it has a bounded time-average backlog"*, i.e., $\limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{E}[Q(t)] \leq \infty$ [16], which can be explained as

$$\text{C4:} \quad \bar{\alpha}_{n_g} \leq \bar{R}_{n_g},$$

where $\bar{\alpha}_{n_g}$ is the average packet arrival rate over one transmission frame. Thus, with the objective to maximize the average rate of users over all slices, the optimization problem can be formulated as

$$\max_{\mathbf{P},\mathbf{w}} \sum_{g \in \mathcal{G}} \sum_{n_g \in \mathcal{N}_g} \bar{R}_{n_g}, \qquad (2)$$

$$\text{subject to C1 - C4,}$$

where $\mathbf{P}$ and $\mathbf{w}$ are the vectors of all $\mathbf{P}_{n_g}$ and $\mathbf{w}_{n_g}$ for all $n_g \in \mathcal{N}_g$ and $g \in \mathcal{G}$, respectively.

Note that (2) is a cross-layer optimization problem which aims to maximize the average rate of system, while considering the queue dynamics of MAC layer and CSI of physical layer. The multidimensional nature and the discrete variables make this problem hard to be solved in this form. In the next section, we resolve these issues via variable relaxation and Lyapunov drift.

III. Solution for Dynamic Resource Provisioning Problem

In the first step, by relaxing the integer variable $w_{n_g,k}^t$, we transform the constraint C2 to

$$\widetilde{\text{C2}} : \sum_{g \in \mathcal{G}} \sum_{n_g \in \mathcal{N}_g} w_{n_g,k}^t \leq 1 \text{ and } w_{n_g,k}^t \in [0,1], \forall k \in \mathcal{K}.$$

In the context of new definition, $w_{n_g,k}^t$ represents the portion of a time slot t for which sub-carrier k is assigned to user n_g. Since by definition, $R_{n_g}^t(\mathbf{P}_{n_g}, \mathbf{w}_{n_g})$ is not convex w.r.t. $\mathbf{P}_{n_g}$ and $\mathbf{w}_{n_g}$, we consider another variable transformation i.e., $x_{n_g,k}^t = w_{n_g,k}^t P_{n_g,k}^t$ to change $R_{n_g}^t(\mathbf{P}, \mathbf{w})$ into a convex function

$$\widetilde{R}_{n_g}^t(\mathbf{x}, \mathbf{w}) = \sum_{k \in \mathcal{K}} w_{n_g,k}^t \log_2 \left(1 + \frac{x_{n_g,k}^t h_{n_g,k}^t}{w_{n_g,k}^t \sigma} \right),$$

similar to the general class of functions $f(x,y) = y \log(1 + x/y)$ [15]. The problem (2), under the new definitions of variables, can be rewritten as

$$\max_{\mathbf{x},\mathbf{w}} \sum_{g \in \mathcal{G}} \sum_{n_g \in \mathcal{N}_g} \bar{\widetilde{R}}_{n_g} \qquad (3)$$

$$\text{subject to } \widetilde{\text{C1}}, \ \widetilde{\text{C2}}, \ \widetilde{\text{C3}}, \text{ and } \widetilde{\text{C4}},$$

where $\bar{\widetilde{R}}_{n_g} = \lim_{T \to \infty} \frac{1}{T} \sum_{t \in T} \widetilde{R}_{n_g}^t(\mathbf{x}, \mathbf{w})$ is the time average rate with new definitions of variables,

$$\widetilde{\text{C1}} : \sum_{g \in \mathcal{G}} \sum_{n_g \in \mathcal{N}_g} \sum_{k \in \mathcal{K}} x_{n_g,k}^t \leq P_{\max},$$

$$\widetilde{C3}: \sum_{n_g \in \mathcal{N}_g} \bar{\tilde{R}}_{n_g} \geq \bar{R}_g^{\text{rsv}} \text{ and } \widetilde{C4}: \bar{\alpha}_{n_g} \leq \bar{\tilde{R}}_{n_g},$$

for all $n_g \in \mathcal{N}_g$ and $g \in \mathcal{G}$.

Since (3) belongs to general stochastic optimization problem, we apply a simple and elegant theory of Lyapunov drift optimization frameworks to convert the average based optimization problem into the single time slot optimization problem which can be solved based on the instantaneous values of CSI, QSI and arrival rate of each user, resulting in much less computational complexity [16]. From the quadratic Lyapunov function, the optimization problem in (3) can be converted into the following optimization problem

$$\min_{\mathbf{x},\mathbf{w}} -V(\sum_{g \in \mathcal{G}} \sum_{n_g \in \mathcal{N}_g} \tilde{R}_{n_g}^t) + \sum_{g \in \mathcal{G}} Z_g^t (\bar{R}_g^{\text{rsv}} - \sum_{n_g \in \mathcal{N}_g} \tilde{R}_{n_g}^t) \quad (4)$$
$$+ \sum_{g \in \mathcal{G}} \sum_{n_g \in \mathcal{N}_g} Q_{n_g}^t (\alpha_{n_g}^t - \tilde{R}_{n_g}^t)$$

subject to : $\widetilde{C1}$ - $\widetilde{C2}$,

where $V > 0$ and Z_g^t are trade-off parameter and virtual queue for constraint $\widetilde{C3}$, respectively in Lyapunov drift-plus-penalty algorithm. The algorithm to solve (4) is presented in Table I. This algorithm contains two major steps: 1) update the queue and virtual queue related parameters, i.e., $Q_{n_g}^t$ and Z_g^t, based on Q1 and Q2 in Table I, respectively; 2) update the optimal power and sub-carrier allocation vectors.

To derive the optimal power and sub-carrier allocation, we can apply the Lagrange dual decomposition to (4) because of its convex nature. The Lagrange function related to (4) is

$$\mathcal{L}(\mathbf{x},\mathbf{w},\lambda,\rho) = -V(\sum_{g \in \mathcal{G}} \sum_{n_g \in \mathcal{N}_g} \tilde{R}_{n_g}^t) + \sum_{g \in \mathcal{G}} Z_g^t \quad (5)$$
$$(\bar{R}_g^{\text{rsv}} - \sum_{n_g \in \mathcal{N}_g} \tilde{R}_{n_g}^t) + \sum_{g \in \mathcal{G}} \sum_{n_g \in \mathcal{N}_g} Q_{n_g}^t (\alpha_{n_g}^t - \tilde{R}_{n_g}^t) + \sum_{k \in \mathcal{K}} \rho_k$$
$$(\sum_{g \in \mathcal{G}_g} \sum_{n_g \in \mathcal{N}_g} w_{n_g,k}^t - 1) + \lambda(P_{\max} - \sum_{g \in \mathcal{G}} \sum_{n_g \in \mathcal{N}_g} \sum_{k \in \mathcal{K}} x_{n_g,k}^t)$$

From KKT conditions related to (4), we obtain the following solution for optimal power allocation

$$P_{n_g,k}^* = \left[\frac{V + Z_g^t + Q_{n_g}^t}{\lambda \ln(2)} - \frac{\sigma^2}{h_{n_g,k}} \right]_0^{P_{\max}}. \quad (6)$$

For sub-carriers, we obtain the following necessary condition for $w_{n_g,k}$, [17]

$$\frac{\partial \mathcal{L}(\mathbf{x},\mathbf{w},\lambda,\rho)}{\partial w_{n_g,k}^{t*}} \begin{cases} < 0, & w_{n_g,k}^{t*} = 0 \\ = 0, & 0 < w_{n_g,k}^{t*} < 1 \quad \forall k \in \mathcal{K}, n_g \in \mathcal{N}_g, \\ > 0, & w_{n_g,k}^{t*} = 1 \end{cases}$$

where $\frac{\partial \mathcal{L}(\mathbf{x},\mathbf{w},\lambda,\rho)}{\partial w_{n_g,k}^{t*}} = \rho_k - (V + Z_g^t + Q_{n_g}^t) \times$

$$\left(\log_2(1 + \gamma_{n_g,k}^t) - \frac{\gamma_{n_g,k}^t}{(1 + \gamma_{n_g,k}^t)\ln(2)} \right), \forall g \in \mathcal{G}. \quad (7)$$

TABLE I
DYNAMIC RESOURCE PROVISIONING ALGORITHM

Algorithm 1 QoS-aware Resource Provisioning Algorithm
1: **Initialization:** Initialize V, $\mathbf{Z} = [Z_1^t, \cdots, Z_G^t]$ and $\mathbf{Q} = [Q_{n_g}]_{n_g \in \mathcal{N}_g}$ for all $g \in \mathcal{G}$, $T_{\max}$ and $i_{\max}$, $\varepsilon = 10^{-3}$.
2: **Step 1:** For $t = 1, \cdots, T_{\max}$
3:      Set $\mathbf{w}^*(i = 0) = 1$, $\mathbf{P}_{n_g}^*(i = 0) = P_{\max}/K$, $\forall n_g \in \mathcal{N}_g, g \in \mathcal{G}$.
4:      Set arrival rate vector $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_{n_g}]_{n_g=1}^{N_g} \; \forall g \in \mathcal{G}$, according to a Poisson random variable.
5:      **Step 2:** Single shot loop iteration i:
6:        $\lambda(i) = \left[\lambda(i-1) + \delta_\lambda \frac{\partial \mathcal{L}}{\partial \lambda}\right]^+$,
7:        Update power and sub-carrier according to (6) and (8),
8:        Select $n_g, k = \max[s_{n_g,k}^t]$ and set $w_{n_g,k}^t = 1$, otherwise $w_{n_g,k}^t = 0$.
9:        When $\|(\lambda(i) - \lambda(i-1))\| \leq \varepsilon$, or $i > i_{\max}$, stop, otherwise go to **Step 2**.
10:    Update:
11:    Q1: $Q_{n_g}^{t+1} = \max[(Q_{n_g}^t - R_{n_g}^t T_s), 0] + \alpha_{n_g}^t$, $\forall n_g \in \mathcal{N}_g, \forall g \in \mathcal{G}$.
12:    Q2: $Z_g^{t+1} = \max[Z_g^t + \bar{R}_g^{\text{rsv}} - \sum_{n_g \in \mathcal{N}_g} R_{n_g}^t, 0]$, $\forall g \in \mathcal{G}$.
13:    When $t \geq T_{\max}$, Stop,
14: Otherwise, set $t = t + 1$, goto **Step 1**.

Therefore, we reach to

$$s_{n_g,k}^t = (V + Z_g^t + Q_{n_g}^t) \times \quad (8)$$
$$\left(\log_2(1 + \gamma_{n_g,k}^t) - \frac{\gamma_{n_g,k}^t}{(1 + \gamma_{n_g,k}^t)\ln(2)} \right), \forall g \in \mathcal{G},$$

where $\gamma_{n_g,k}^t = \frac{x_{n_g,k}^t h_{n_g,k}^t}{\sigma w_{n_g,k}^t}$. The sub-carrier k is allocated to user n_g, i.e., $w_{n_g,k}^t = 1$ for which $s_{n_g,k}^t$ is maximum otherwise $w_{n_g,k}^t = 0$. In the inner loop of algorithm in Table I, the power and sub-carriers are allocated according to current CSI in the time slot t for fixed queues, i.e., Q_{n_g} for all users and Z_g for all slices. Then, Q_{n_g} and Z_g are updated according to the allocated power and sub-carriers in the outer loop until $t = T_{\max}$ where $T_{\max}$ is a maximum number of time slots.

Via this algorithm, the solution for time-average-based problem (3) is derived from (4) which is instantaneously solved at each time slot based on the CSI values of this slot and without considering the average over other time slots. Now, the important question is to declare the relationship between solutions of (3) and (4). This relation is discussed in the following theorem.

Theorem 1: The Lyapunov quadratic approximation of (3) satisfies all the constraints in (2) and the obtained rate is less than R_{opt} by a gap of B/V, which decreases with increasing V and $B > 0$ derived based on [6] .

Proof. Due to power limitation, for this setup, there exist two

finite constants B_1 and B_2 which satisfy

$$\frac{1}{2}\sum_{g\in\mathcal{G}}\mathbb{E}\{\sum_{n_g\in\mathcal{N}_g}\widetilde{R}_{n_g}^t-\bar{R}_g^{\mathrm{rsv}}\}^2\leq B_1,$$

$$\frac{1}{2}\sum_{g\in\mathcal{G}}\sum_{n_g\in\mathcal{N}_g}\mathbb{E}\{\bar{\alpha}_{n_g}-\widetilde{R}_{n_g}^t\}^2\leq B_2.$$

Assuming $V\geq 0$ and $Z_g^1=0,\forall g\in\mathcal{G}$, all the desired constraints in (3) are satisfied, i.e.,

$$\limsup_{T\to\infty}\frac{1}{T}\sum_{t\in\mathcal{T}}\mathbb{E}\{\sum_{n_g\in\mathcal{N}_g}\widetilde{R}_{n_g}^t-\bar{R}_g^{\mathrm{rsv}}\}\leq 0,\forall g\in\mathcal{G},$$

$$\limsup_{T\to\infty}\frac{1}{T}\sum_{t\in\mathcal{T}}\mathbb{E}\{\bar{\alpha}_{n_g}-\widetilde{R}_{n_g}^t\}\leq 0,\forall n_g\in\mathcal{N}_g,\ \forall g\in\mathcal{G}.$$

Also, if R_{opt} is the maximum optimal rate, then the average expectation of objective function in (3) satisfies

$$\frac{1}{T}\sum_{t\in\mathcal{T}}\mathbb{E}\{\sum_{g\in\mathcal{G}}\sum_{n_g\in\mathcal{N}_g}\widetilde{R}_{n_g}^t\}\leq R_{\mathrm{opt}}+\frac{B}{V},\quad\forall t\in\mathcal{T},$$

where $B=B_1+B_2$. Furthermore, for all time slots $t\in\mathcal{T}$, we have

$$\frac{1}{T}\sum_{t\in\mathcal{T}}\mathbb{E}\{\sum_{n_g\in\mathcal{N}_g}\widetilde{R}_{n_g}^t-\widetilde{R}_g^{\mathrm{rsv}}\}\tag{9a}$$

$$\leq\mathbb{E}\{Z_g[T+1]/T\},\forall g\in\mathcal{G},$$

$$\frac{1}{T}\sum_{t\in\mathcal{T}}\mathbb{E}\{\bar{\alpha}_{n_g}-\widetilde{R}_{n_g}^t\}\tag{9b}$$

$$\leq\mathbb{E}\{Q_{n_g}[T+1]/T\},\forall n_g\in\mathcal{N}_g,\ \forall g\in\mathcal{G}$$

and

$$\sum_{g\in\mathcal{G}}\mathbb{E}\{Z_g[T+1]^2/T^2\}+\sum_{g\in\mathcal{G}}\sum_{n_g\in\mathcal{N}_g}\mathbb{E}\{Q_{n_g}[T+1]^2/T^2\}$$

$$\leq\frac{1}{T}[2B+2V(R_{\mathrm{opt}}-\theta)],\tag{10}$$

where θ is a finite constant such that $\mathbb{E}\{\sum_{g\in\mathcal{G}}\sum_{n_g\in\mathcal{N}_g}\widetilde{R}_{n_g}^t\}\geq\theta,\forall t\in\mathcal{T}$. Such θ exists because $\widetilde{R}_{n_g}^t$ is lower bounded [6]. $\square$

The above statement declares that via the Lyapunov drift approach, when (2) is feasible, the solution of (4) satisfies the constraints of (2), and it is within $O(1/V)$ of the optimum value of (2). Therefore, the time-average based solution can be derived by solution of optimization problem in each frame. The difference between solutions of (2) and (4), i.e., $\frac{B}{V}$, can be made arbitrarily small by increasing the value of parameter V. However, increasing the value of V affects the convergence time, as shown by (9) and (10). Consequently, V is a trade-off parameter for convergence time and optimality of the proposed algorithm in Table I [6].

IV. NUMERICAL RESULTS AND DISCUSSIONS

The simulation setup consists of one BS with $K=128$ sub-carriers which serves $G=3$ slices. Each slice $g\in\mathcal{G}$ contains
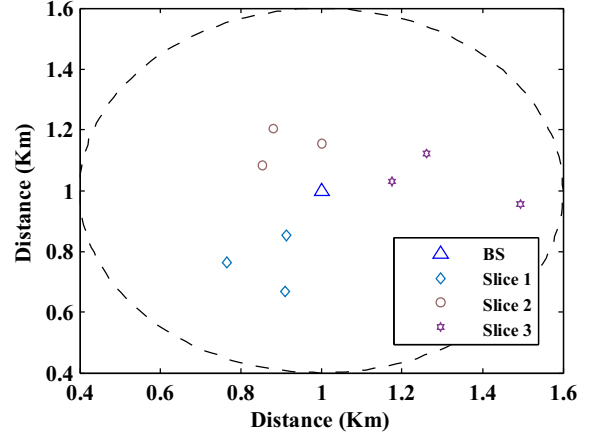

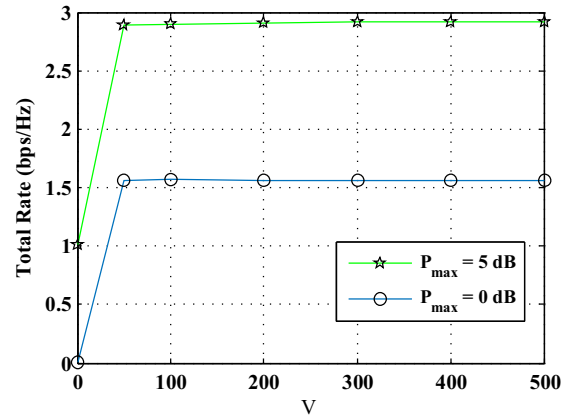Fig. 1. Locations of users of three slices.

Fig. 2. Total Rate versus V and $P_{\max}$.

$N_g=3$ users. The locations of all the users from different slices are fixed in the range of distance $d_{n_g}\in\{0.1,\ 0.6\}$ Km as shown in the Fig. 1. Furthermore, under Rayleigh fading assumption, the channel power gain of user n_g on sub-carrier k is modeled as $h_{n_g,k}=\mathcal{X}_{n_g,k}/d_{n_g}^\beta$, where $\beta=3$ is the path-loss exponent and $\mathcal{X}_{n_g,k}$ is the exponential random variable for user n_g on sub-carrier k with mean one. We set $P_{\max}=5$ dB, $\sigma=1$, $\bar{R}_g^{\mathrm{rsv}}=0.5$ bps/Hz, and average Poisson packet arrival rate $\lambda=1$ packet/slot with each packet of size $l=1$ bits/Hz/slot, unless otherwise stated. All the simulation results are generated with $T_{\max}=5000$ time slots with each time slot length $t=1$ ms for the bandwidth $B_c=200$ KHz.

Fig. 2 shows the total rate versus V for various values of $P_{\max}$. It is observed that the total rate is increased with increasing V. It is expected because the proposed optimization problem (2) is solved via Lyapunov drift-plus-penalty approach, which gives asymptotically optimal solution in terms of average rate over time as V approaches infinity [6]. In other words, the gap between the obtained average rate and optimal average rate reduces to zero with increasing V. Next, we study the effect of packet size arriving in users' queue on average total rate. Fig.3 illustrates the decrease in total rate with increasing packet size l. This is because the feasibility region shrinks with increasing l due to constraint
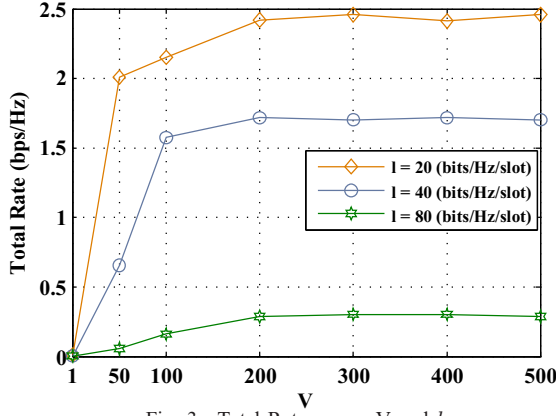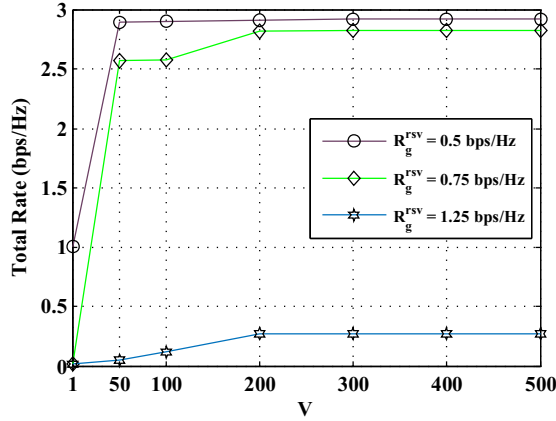
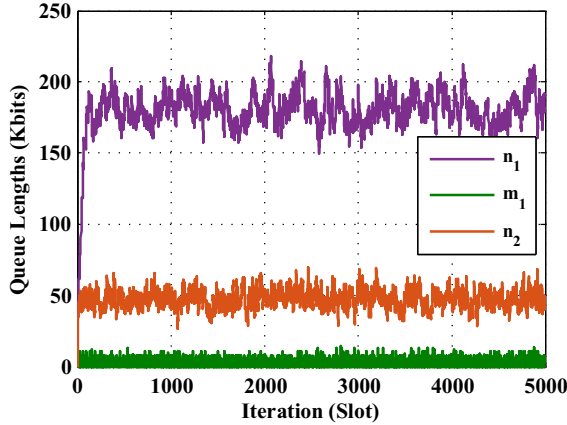Fig. 3. Total Rate versus V and l.

Fig. 4. Total Rate versus V and R_g^{rsv}.

Fig. 5. Queue Lengths versus Slot iterations.

words, network queue stability condition holds for the WVN, confirming the effectiveness of the proposed algorithm.

V. CONCLUSION

In this paper, we proposed a QoS-oriented dynamic resource provisioning policy for OFDMA-based WVNs which simultaneously satisfies the condition of stable average queue lengths of the users of a slice and the minimum reserved rate of each slice. The solution is proposed by transforming the cross-layer problem into physical layer problem via Lyapunov drift-plus-penalty approach. Finally, the simulation results under different system parameters confirm the effectiveness of the proposed algorithm.

REFERENCES

[1] R. Kokku, R. Mahindra, H. Zhang, and S. Rangarajan, "NVS: A substrate for virtualizing wireless resources in cellular networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 5, pp. 1333–1346, Oct. 2012.

[2] F. Fu and U. Kozat, "Stochastic game for wireless network virtualization," *IEEE/ACM Trans. Netw.*, vol. 21, no. 1, pp. 84–97, Feb. 2013.

[3] S. Parsaeefard, V. Jumba, M. Derakhshani, and T. Le-Ngoc, "Joint resource provisioning and admission control in wireless virtualized networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, March 2015, pp. 2020–2025.

[4] V. Jumba, S. Parsaeefard, M. Derakhshani, and T. Le-Ngoc, "Resource provisioning in wireless virtualized networks via massive-mimo," *IEEE Wireless Commun. Letters*, 2015.

[5] X. Zhang, Y. Li, D. Jin, L. Su, L. Zeng, and P. Hui, "Efficient resource allocation for wireless virtualization using time-space division," in *Intl. Wireless Commun. and Mobile Computing Conf. (IWCMC)*, Aug. 2012.

[6] B. Liu, P. Terlecky, A. Bar-Noy, R. Govindan, M. Neely, and D. Rawitz, "Optimizing information credibility in social swarming applications," *IEEE Trans. Parallel and Distributed Systems*, vol. 23, no. 6, pp. 1147–1158, Jun. 2012.

[7] H. Wang and Z. Ding, "Macrocell-queue-stabilization-based power control of femtocell networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5223–5236, Sep. 2014.

[8] M. Neely, E. Modiano, and C. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 89–103, January 2005.

[9] M. Neely, "Optimal peer-to-peer scheduling for mobile wireless networks with redundantly distributed data," *IEEE Transactions on Mobile Computing*, vol. 13, no. 9, pp. 2086–2099, September 2014.

[10] N. M.J., "Delay-based network utility maximization," *IEEE/ACM Trans. Netw.*, vol. 21, no. 1, pp. 41–54, Feb. 2013.

[11] L. Xie and X. Zhang, "TDMA and FDMA based resource allocations for quality of service provisioning over wireless relay networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2007.

[12] R. Urgaonkar and M. Neely, "Opportunistic cooperation in cognitive femtocell networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 607–616, April 2012.

[13] M. Lessinnes, J.-M. Dricot, P. De Doncker, L. Vandendorpe, and F. Horlin, "Dynamic resource allocation for MIMO cognitive networks with low control traffic and low computational complexity," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1732–1740, May 2013.

[14] W. Cheng, X. Zhang, and H. Zhang, "Optimal dynamic power control for full-duplex bidirectional-channel based wireless networks," in *Proc. IEEE Intl. Conf. on Computer Commun. (INFOCOM)*, Apr. 2013.

[15] D. Ng, E. Lo, and R. Schober, "Energy-efficient resource allocation in ofdma systems with large numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3292–3304, Sep. 2012.

[16] Y. Cui, V. Lau, R. Wang, H. Huang, and S. Zhang, "A survey on delay-aware resource control for wireless systems;large deviation theory, stochastic lyapunov drift, and distributed stochastic learning," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1677–1701, Mar. 2012.

[17] M. Tao, Y.-C. Liang, and F. Zhang, "Resource allocation for delay differentiated traffic in multiuser OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2190–2201, Jun. 2008.

[18] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2009.

C4 in problem (2) which decreases the obtained achievable rate [18]. Furthermore, since increasing R_g^{rsv} also shrinks the feasibility region due to constraint C3, the total rate decreases with increasing R_g^{rsv} as depicted in Fig.4. However, the total average rate increases with increasing V for both Fig.3 and Fig.4, which further confirms our observations in Fig.2.

To further demonstrate the convergence of Alg. 1 in terms of queue stability, in Fig.5, we plot the queue lengths of farthest users n_1 and m_1 from slice 1 in addition to user n_2 from slice 2 at each time slot. It is observed that the queue lengths of all users remains bounded as time slots increases. In other