

ENERGY-EFFICIENCY VERSUS DELAY TRADEOFF IN WIRELESS NETWORKS VIRTUALIZATION

Qiong Shi, Liqiang Zhao, *Member, IEEE*, Yaoyuan Zhang, Gan Zheng, *Senior Member, IEEE*,
F. Richard Yu, *Senior Member, IEEE*, and Hsiao-Hwa Chen *Fellow, IEEE*

[†]Corresponding Author's Address:

Hsiao-Hwa Chen

Department of Engineering Science

National Cheng Kung University

1 Da-Hsueh Road, Tainan City, 70101, Taiwan, ROC

Tel: +886-6-2757575 ext. 63320

Fax: +886-6-2766549

Email: hshwchen@ieee.org

Qiong Shi (e-mail: q.shi@outlook.com), Liqiang Zhao (e-mail: lqzhao@mail.xidian.edu.cn), and Yaoyuan Zhang (e-mail: yyzhang@s-an.org) are with the State Key Laboratory of Integrated Service Networks, Xidian University, China. Gan Zheng (e-mail: g.zheng@lboro.ac.uk) is with the Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, UK. F. R. Yu (e-mail: Richard.yu@carleton.ca) is with the Department of System and Computer Engineering, Carleton University, Canada. Hsiao-Hwa Chen (e-mail: hshwchen@ieee.org) is with the Department of Engineering Science, National Cheng Kung University, Taiwan.

The paper was submitted July 31, 2017.

Energy-efficiency versus Delay Tradeoff in Wireless Networks Virtualization

Qiong Shi, Liqiang Zhao, *Member, IEEE*, Yaoyuan Zhang, Gan Zheng, *Senior Member, IEEE*,
F. Richard Yu, *Senior Member, IEEE*, and Hsiao-Hwa Chen *Fellow, IEEE*

Abstract—This paper studies the issues on wireless networks virtualization in terms of two important performance metrics, i.e., energy efficiency (EE) and delay. Different from existing works on physical layer, we aim to achieve a good tradeoff between EE and delay in wireless networks virtualization using cross-layer stochastic optimization approach. In particular, we formulate a cross-layer problem using fractional programming and Lyapunov optimization method. The EE and delay tradeoff solution is given explicitly by deriving their analytical bounds that are verified by simulation results.

Index Terms—Wireless network virtualization; Energy efficiency; Delay

I. INTRODUCTION

To support rapidly growing high-speed data traffic with satisfactory user experience and to reduce the cost and global carbon dioxide emission, EE has become one of the major design goals in 5G systems [1]. Delay is another critical performance indicator and is related directly to system reliability. [2]. The EE-delay tradeoff was observed by simulations in orthogonal frequency division multiple access (OFDMA) systems and multiple-input-multiple-output (MIMO) systems [3]. A QoS model for wireless networks virtualization (WNV) was developed in [5]. The optimal resource allocation was studied for WNV in [6]. To the best of our knowledge, the EE and EE-delay tradeoff in WNV have not been well investigated in the literature, which is the focus of this work.

In this paper, first we will formulate an EE-delay tradeoff problem in WNV, in which data arrive randomly at base stations (BSs) in a physical substrate network and are queued for transmission. We model our problem as a stochastic optimization to maximize EE while guarantee finite virtual user queue length. In traditional wireless networks, data arrive and are dealt with independently, and the performance has been studied in various network models with a finite queue length [14]–[16]. In our model, virtual BSs cooperate to provide service to the users through the WNV. Actual queues are therefore transformed into several virtual user queues according to user customized requirements. Our virtual user

queue representation is a result of network virtualization, which has not been studied in the literature, and existing methods cannot be applied to solve this problem directly.

II. SYSTEM MODEL

Let us consider a one-hop wireless communication network consisting of K users and M cells (each cell has one physical BS). N virtual cells (or N virtual BSs) can be formed through abstracting, slicing, isolating, and sharing wireless network infrastructure as well as radio spectrum resources belonging to the M cells, similar to the scenarios as shown in Fig. 4 of [1].

Assume that the system is time slotted and slot t is normalized to an integer unit described as a time interval $[t, t + 1)$, where $t \in \{0, 1, 2, \dots\}$. Data arrive randomly in every slot at each BS and queue for being served. After virtualization, actual queuing process in BSs can be transformed into N virtual queues, whose data are sent logically by a WNV. Furthermore, these queues can be coalesced into K virtual user queues to serve K users respectively, as depicted in Fig. 1. Let vectors $A(t) = \{A_1(t), A_2(t), \dots, A_K(t)\}$ and $Q(t) = \{Q_1(t), Q_2(t), \dots, Q_K(t)\}$ denote the processes of random data arrivals and the current queue lengths in K virtual user queues at slot t , respectively, where $A(t)$ is independent and identically distributed (i.i.d.) over time and its arrival rate is λ .

The channel conditions between virtual BS n and user k are described by $S_{n,k}(t)$, which is i.i.d. over time. $S_{n,k}(t)$ remains unchanged in a slot, and may change from slot to slot. Similarly, $P_{n,k}(t)$ represents transmit power from virtual BS n to user k . Furthermore, if a continuous rate is adopted and ϱ^2 is the noise power, the transmit power of a virtual BS n and achievable rate of user k from the WNV¹ can be expressed as

$$P_n(t) = \sum_{k=1}^K P_{n,k}(t) + P_n^c, \quad (1)$$

$$R_k[P(t), S(t)] = \sum_{n=1}^N \log_2 \left[1 + \frac{P_{n,k}(t)S_{n,k}(t)}{\varrho^2} \right], \quad (2)$$

respectively, where $S(t)$ and $P(t)$ are metrics of channel conditions and power allocation, respectively, which can be expressed as $P(t) = (P_{n,k}(t))$ and $S(t) = (S_{n,k}(t))$. P_n^c denotes the power consumption of the circuits.

¹Since the virtual BSs slices are generated from virtualization and isolated from each other, we do not consider the interference among them.

Qiong Shi (e-mail: q.shi@outlook.com), Liqiang Zhao (e-mail: lqzhao@mail.xidian.edu.cn), and Yaoyuan Zhang (e-mail: yyzhang@s-an.org) are with the State Key Laboratory of Integrated Service Networks, Xidian University, China. Gan Zheng (e-mail: g.zheng@lboro.ac.uk) is with the Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, UK. F. R. Yu (e-mail: Richard.yu@carleton.ca) is with the Department of System and Computer Engineering, Carleton University, Canada. Hsiao-Hwa Chen (e-mail: hshwchen@ieee.org) is with the Department of Engineering Science, National Cheng Kung University, Taiwan.

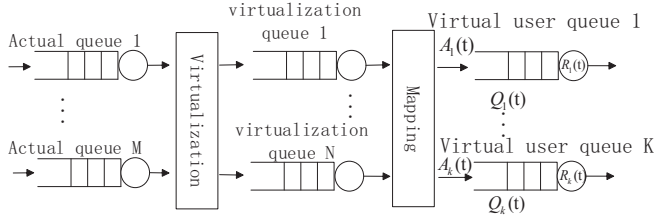


Fig. 1. The actual queuing process and virtual queue representations.

Meanwhile, we can have the time-average expectation of R_k as

$$\bar{R}_k(P, S) = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} E\{R_k[P(\tau), S(\tau)]\}, \quad (3)$$

where P and S are defined as $P = \{P(0), P(1), \dots\}$ and $S = \{S(0), S(1), \dots\}$, respectively. Based on (2) and (2), the power consumption and transmit rate of the WNV are given by

$$P_{wvn}(t) = P_{wvn}[P(t)] = \sum_{n=1}^N P_n(t), \quad (4)$$

$$R_{wvn}(t) = R_{wvn}[P(t), S(t)] = \sum_{k=1}^K R_k[P(t), S(t)]. \quad (5)$$

Similar to (3), the time averages of P_{wvn} and R_{wvn} can be defined as

$$\bar{P}_{wvn}(P) = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} E[P_{wvn}(\tau)], \quad (6)$$

$$\bar{R}_{wvn}(P, S) = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} E\{R_{wvn}[P(\tau), S(\tau)]\} \quad (7)$$

respectively, and $E[\cdot]$ denotes the expectation. Then, we can define the long-term EE as

$$U_{EE} = \frac{\bar{R}_{wvn}(P, S)}{\bar{P}_{wvn}(P)}. \quad (8)$$

Next, we derive the delay metric. Based on Fig. ??, we can model the queuing process as

$$Q_k(t+1) = \max[Q_k(t) - R_k(t), 0] + A_k(t), \quad \forall k. \quad (9)$$

Note that users can be served according to their customized requirements through virtualization networks, such as delay-tolerant and delay-sensitive services, which lead to different QoS requirements among users. Therefore, we define a parameter $D_k(t)$, called equivalent queue length, to measure the queue lengths in a network, which is defined as

$$D_k(t) = \frac{Q_k(t)}{\gamma_k}, \quad (10)$$

where γ_k is the equivalent queue factor that is set according to users' diverse QoS requirements.

In this paper, we focus on the problem of EE-delay tradeoff in a steady-state network. A network is steady if all discrete time processes $Q_k(t)$ are mean rate stable, that is, satisfying the following condition [7]

$$\lim_{t \rightarrow \infty} \frac{E\{|Q_k(t)|\}}{t} = 0. \quad (11)$$

III. PROBLEM FORMULATION AND SOLUTION

In this section, we investigate the EE-delay tradeoff in WVN by solving the EE maximization problem with the constraints on user rate requests R_k^{av} and delay limit β . Mathematically, the problem can be formulated as

$$\begin{aligned} \max_P \quad & U_{EE} = \frac{\bar{R}_{wvn}(P, S)}{\bar{P}_{wvn}(P)}, \\ \text{s.t.} \quad & C1: \bar{R}_k \geq R_k^{av}, \quad \forall k, \\ & C2: D_k(t) \leq \beta, \quad \forall k, t, \\ & C3: P_{n,k}(t) \geq 0, \quad \forall n, t, \\ & C4: P_{wvn}(t) \leq \sum_{m=1}^M P_m^{max}, \quad \forall m, t. \end{aligned} \quad (12)$$

It can be seen that $C1$ is to satisfy the average rate R_k^{av} requested by each user. $C2$ guarantees the stability of queues and ensures that the equivalent queue length is below a threshold β , which can be set according to user requirements. Moreover, $C2$ also gives the delay constraint by Little's Theorem. $C3$ gives a non-negative power constraint, and $C4$ limits the maximum power of the WNV when the maximum transmit power of physical BS m is P_m^{max} .

Due to the nonlinear fraction in the objective function, we cannot solve the problem (12) directly. For simplicity, let $X_{EE} = \frac{1}{U_{EE}}$. It is easy to see that $\max_P U_{EE}$ is equivalent to $\min_P X_{EE}$. Moreover, define a feasible set of (12) as ψ , and let P^{opt} be the optimal power allocation. Then, the optimal solution to achieve both U_{EE}^{opt} and X_{EE}^{opt} becomes

$$X_{EE}^{opt} = \frac{1}{U_{EE}^{opt}} = \frac{\bar{P}_{wvn}(P^{opt})}{\bar{R}_{wvn}(P^{opt}, S)} = \min_{P \in \psi} \frac{\bar{P}_{wvn}(P)}{\bar{R}_{wvn}(P, S)}. \quad (13)$$

According to the generalized fractional programming theory [8], the optimal U_{EE}^{opt} can be achieved if and only if

$$\begin{aligned} \min_{P \in \psi} \quad & \bar{P}_{wvn}(P) - X_{EE}^{opt} \bar{R}_{wvn}(P, S), \\ & = \bar{P}_{wvn}(P^{opt}) - X_{EE}^{opt} \bar{R}_{wvn}(P^{opt}, S) = 0. \end{aligned} \quad (14)$$

The original problem (12) is then transformed to the following equivalent problem:

$$\begin{aligned} \min \quad & \bar{P}_{wvn}(P) - X_{EE}^{opt} \bar{R}_{wvn}(P, S), \\ \text{s.t.} \quad & C1 - C4. \end{aligned} \quad (15)$$

The existence of time average expectations and $C2$ in problem (15) determines that Dinkelbachs algorithm developed in [8] can not be used. Thus, In order to develop an effective algorithm based on classical drift-plus-penalty algorithm, we introduce $X_{EE}(t)$, where $X_{EE}(0) = 0$ and $t \in \{1, 2, \dots\}$, depending on the past power allocation, or

$$X_{EE}(t) = \frac{\sum_{\tau=0}^{t-1} P_{wvn}[P(\tau)]}{\sum_{\tau=0}^{t-1} R_{wvn}[P(\tau), S(\tau)]}. \quad (16)$$

The optimal solution of the original problem (12) can be obtained by replacing X_{EE}^{opt} by $X_{EE}(t)$ in (15) and minimizing the following problem sequentially in each time slot, or

$$\begin{aligned} \min \quad & \bar{P}_{wvn}(P) - X_{EE}(t) \bar{R}_{wvn}(P, S), \\ \text{s.t.} \quad & C1 - C4. \end{aligned} \quad (17)$$

Next, we show how to simplify $C1$ and $C2$. Based on the general Lyapunov theory [7], $C1$ in (12) can be transformed into a queue stable problem by introducing virtual rate queues $G_k(t)$ ($\forall k, t$), where $G_k(0) = 0$ and

$$G_k(t+1) = \max [G_k(t) + R_k^{av} - R_k(t), 0]. \quad (18)$$

Combining (9) and (10), we can see that constraint $C2$ can be written equivalently as the following constraint $\tilde{C}2$:

$$Q_k(t+1) = \begin{cases} \max [Q_k(t) - R_k(t), 0], & \text{if } D_k(t) > \beta, \\ \max [Q_k(t) - R_k(t), 0] + A_k(t), & \text{e.w.} \end{cases} \quad (19)$$

Accordingly, (17) can be further recast equivalently to

$$\min \bar{P}_{wvn}(P) - X_{EE}(t)\bar{R}_{wvn}(P, S), \quad (20)$$

s.t. $\tilde{C}2, C3, C4$,

$C1 : G_k(t)$ is mean rate stable, $\forall k$.

To tackle $C1$ and $C2$, let $\Theta(t) = [Q(t), G(t)]$ be a combined vector to represent queuing states of all queues, where $Q(t)$ is a vector of virtual user queues in the network and $G(t)$ is virtual rate queues. Then, the Lyapunov function $L[\Theta(t)]^2$ and conditional Lyapunov drift $\Delta[\Theta(t)]$ can be defined as

$$L(\Theta(t)) = \frac{1}{2} \sum_{k=1}^K Q_k(t)^2 + \frac{1}{2} \sum_{k=1}^K G_k(t)^2, \quad (21)$$

$$\Delta[\Theta(t)] = E\{L[\Theta(t+1)] - L[\Theta(t)] \mid \Theta(t)\}. \quad (22)$$

According to Lyapunov optimization theory [7], the solution of the above problem (20) can be obtained by minimizing the upper bound of the drift-plus-penalty expression $F(t)$ slot by slot as follows

$$F(t) = \Delta[\Theta(t)] - VE[P_{wvn}(t) - X_{EE}(t)R_{wvn}(t)], \quad (23)$$

where $V \geq 0$ is used for controlling the tradeoff between the EE and delay, which will be explained later.

The bound of $F(t)$ can be determined in (24) using Lyapunov optimization theory [10], or

$$\begin{aligned} F(t) &\leq B + \sum_{k=1}^K \xi [D_k(t) - \beta] Q_k(t) E[A_k(t) | \Theta(t)] \\ &\quad - \sum_{k=1}^K Q_k(t) E[R_k(t) | \Theta(t)] + \sum_{k=1}^K G_k(t) E[R_k^{av} \\ &\quad - R_k(t) | \Theta(t)] + VE[P_{wvn}(t) - X_{EE}(t)R_{wvn}(t) | \Theta(t)], \end{aligned} \quad (24)$$

where $B > 0$ is a constant and satisfies the following condition:

$$\begin{aligned} B &\geq \frac{1}{2} \sum_{k=1}^K E\{\xi [D_k(t) - \beta] A_k(t)^2 + R_k(t)^2 | \Theta(t)\} \\ &\quad + \frac{1}{2} \sum_{k=1}^K \{ [R_k^{av} - R_k(t)]^2 | \Theta(t)\}, \end{aligned} \quad (25)$$

where ξ is a function of $D_k(t) - \beta$ and satisfies the following condition: $D_k(t) - \beta \leq 0$, $\xi [D_k(t) - \beta] = 1$; otherwise, $\xi [D_k(t) - \beta] = 0$, which satisfies $\tilde{C}2$ in (19). Thus, the optimal

²Note that other Lyapunov functions could also be used as in [9], and they may lead to a different performance, but this is out of the scope of this paper.

power allocation in slot t can be found by minimizing the right-hand-side of (24), i.e., solving the problem (26) below:

$$\begin{aligned} \min &\sum_{k=1}^K \left\{ G_k(t) [R_k^{av} - R_k(t)] - Q_k(t) [R_k(t)] \right. \\ &\quad \left. + \xi [D_k(t) - \beta] Q_k(t) A_k(t) \right\} \\ &+ V \left[\sum_{n=1}^N P_n(t) - X_{EE}(t) \sum_{k=1}^K R_k(t) \right], \quad (26) \\ \text{s.t. } &C3 : P_{n,k}(t) \geq 0, \forall n, t, \\ &C4 : P_{wvn}(t) \leq \sum_{m=1}^M P_m^{max}, \forall m, t. \end{aligned}$$

It is easy to see that (26) is a convex problem and its optimal solution follows a standard water-filling structure [10]. Thus, the complexity of solving it is $O(W^2)$ [11], where $W = \max(N, K)$. The overall algorithm to solve (12) is summarized in Algorithm 1 and the complexity is $O(TW^2)$. The overall algorithm to solve (12) is summarized in Algorithm 1 and the complexity is $O(TW^2)$.

Algorithm 1 Dynamic power allocation algorithm to solve (12).

- 1: Initialization: $Q_k(0)=0$, $G_k(0)=0$, and $X_{EE}(0)=0$, $\forall m$.
- 2: **Repeat**:
- 3: Update equivalent queue length $D_k(t)$ according to (10).
- 4: Update power allocate $P(t) = [P_{n,k}(t)]$ as the optimal solution to problem (26).
- 5: $t = t + 1$.
- 6: Update actual queue length $Q_k(t)$ according to (19).
- 7: Update virtual queue length $G_k(t)$ according to (18).
- 8: Update $X_{EE}(t)$ according to (16).
- 9: **Stop** when $t = T$, where T is the total number of time slots.

It is worth mentioning that the proposed algorithm does not necessarily converge to a global solution of the original problem, but only narrows down the gap between the upper bound solution and the optimal solution of the original problem. However, instead of solving the original problem, our main purpose is to analyze the performance of network and derive the theoretical bounds for both EE and delay.

IV. ENERGY EFFICIENCY AND DELAY TRADEOFF

In this section, we will give the analysis on the tradeoff between the EE and delay. Assume that the problem (20) is feasible and $E\{L[\Theta(0)]\}$ is bounded. If λ and $\lambda + \varepsilon$ for a positive ε do not exceed network capacity of the WNV, we have the following properties for (26).

Theorem 1: The EE in (12) is bounded by

$$\frac{U_{EE}^{opt}}{1 + \frac{BU_{EE}^{opt}/R_{min}}{V}} \leq U_{EE} \leq U_{EE}^{opt}, \quad (27)$$

where R_{min} is a finite constant to bound R_{wvn} .

Proof: As shown in [7], [12], $P_{wvn}^*(t)$, $R_{wvn}^*(t)$, and $R_k^*(t)$ are the values acquired by an arbitrary power allocation

policy $P^*(t)$ in slot t , and they satisfy the following conditions for any $\delta > 0$, or

$$E[P_{wvn}^*(t)|\Theta(t)] = E[P_{wvn}^*(t)] \quad (28)$$

$$\leq E[R_{wvn}^*(t)](X_{EE}^{opt} + \delta),$$

$$E[R_{wvn}^*(t)|\Theta(t)] = E[R_{wvn}^*(t)], \quad (29)$$

$$E[R_k^*(t)|\Theta(t)] = E[R_k^*(t)] \geq \lambda_k + \varepsilon, \quad (30)$$

$$E[R_k^{av} - R_k^*(t)|\Theta(t)] = E[R_k^{av} - R_k^*(t)] \leq \delta. \quad (31)$$

Substituting (29)-(31) into the right-hand side of (24) and letting $\delta \rightarrow 0$, we have the following inequality:

$$\begin{aligned} & \Delta [\Theta(t)] + VE[P_{wvn}(t) - X_{EE}(t)R_{wvn}(t)|\Theta_t] \\ & \leq B + \sum_{k=1}^K \{\xi[D_k(t) - \beta] - 1\} Q_k(t)\lambda_k - \varepsilon \sum_{k=1}^K Q_k(t) \quad (32) \\ & + VX_{EE}^{opt}(t)E[R_{wvn}^*(t)] - VX_{EE}(t)E[R_{wvn}^*(t)], \end{aligned}$$

where δ and ε are two constants. Taking expectation while using telescoping sum over $t \in \{0, 1, \dots, H-1\}$ for (32), where H is the number of the time slots, and considering the fact that $Q_k(t) \geq 0$, we have

$$\begin{aligned} & E\{L[\Theta(H)]\} - E\{L[\Theta(0)]\} \\ & + V \left\{ \sum_{t=0}^{H-1} E[P_{wvn}(t)] - \sum_{t=0}^{H-1} E[X_{EE}(t)P_{wvn}(t)] \right\} \\ & \leq HB + HVX_{EE}^{opt}E[R_{wvn}^*(t)] \quad (33) \\ & - VE(R_{wvn}^*) \sum_{t=0}^{H-1} E[X_{EE}(t)]. \end{aligned}$$

Dividing both sides of (33) by VH and making a rearrangement, we get

$$\begin{aligned} & \frac{E\{L[\Theta(0)]\}}{VH} + \frac{1}{H} \sum_{t=0}^{H-1} E[P_{wvn}(t)] - \frac{1}{H} \sum_{t=0}^{H-1} E[X_{EE}(t) \\ & \times R_{wvn}(t)] \leq \frac{B}{V} - E[R_{wvn}^*(t)] \frac{1}{H} \sum_{t=0}^{H-1} E[X_{EE}(t)] \quad (34) \\ & + X_{EE}^{opt}E[R_{wvn}^*(t)]. \end{aligned}$$

Let $H \rightarrow \infty$. The bounds become

$$R_{\min} \leq E\{R_{wvn}[P(t), S(t)]\} \leq R_{\max}, \quad (35)$$

where R_{\max} is a finite constant. Then, we have

$$X_{EE} \leq X_{EE}^{opt} + \frac{B}{VE[R_{wvn}^*(t)]} \leq X_{EE}^{opt} + \frac{B}{VR_{\min}}. \quad (36)$$

Due to the fact that we have $U_{EE}^{opt} = \frac{1}{X_{EE}^{opt}}$, the bounds of EE in (27) are proved. ■

Next, let us study the bounds on the average queue length and the results are summarized in the following theorem.

Theorem 2: The average queue length is bounded by

$$\begin{aligned} \bar{Q} & = \lim_{H \rightarrow \infty} \frac{1}{H} \sum_{t=0}^{H-1} \left\{ \sum_{k=1}^K E[Q_k(t)] \right\} \\ & \leq \min \left[\frac{B + V(R_{\max}/U_{EE}^{opt} + P_{\max})}{\varepsilon}, \sum_{k=1}^K \gamma_k \beta \right], \quad (37) \end{aligned}$$

where P_{\max} is a finite constant to bound P_{wvn} .

Proof: When $D_k(t) \leq \beta$, divide both sides of the inequality (32) by εH , and take $H \rightarrow \infty$. (32) can be further written as

$$\begin{aligned} \bar{Q} & = \lim_{H \rightarrow \infty} \frac{1}{H} \sum_{t=0}^{H-1} \left\{ \sum_{k=1}^K E[Q_k(t)] \right\} \\ & \leq \frac{B + VX_{EE}^{opt}E[R_{wvn}^*(t)]}{\varepsilon} \quad (38) \\ & + \frac{V}{\varepsilon} \lim_{H \rightarrow \infty} \frac{1}{H} \sum_{t=0}^{H-1} E[X_{EE}(t)R_{wvn}(t)], \end{aligned}$$

where P_{wvn} is assumed to be bounded by constants P_{\min} and P_{\max} as

$$P_{\min} \leq E\{P_{wvn}[P(t), S(t)]\} \leq P_{\max}, \quad (39)$$

$$\lim_{H \rightarrow \infty} \frac{1}{H} \sum_{t=0}^{H-1} E[X_{EE}(t)R_{wvn}(t)] \leq P_{\max}, \quad (40)$$

$$E[R_{wvn}^*(t)] \leq R_{\max}. \quad (41)$$

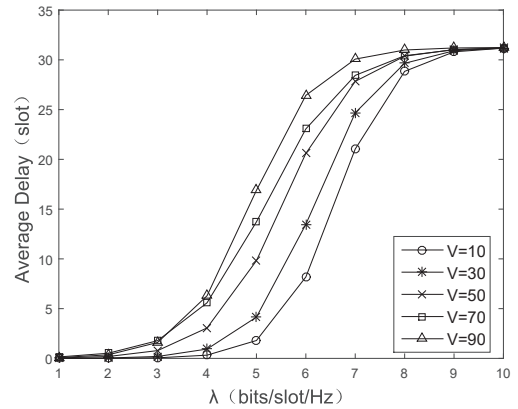


Fig. 2. Average delay versus λ .

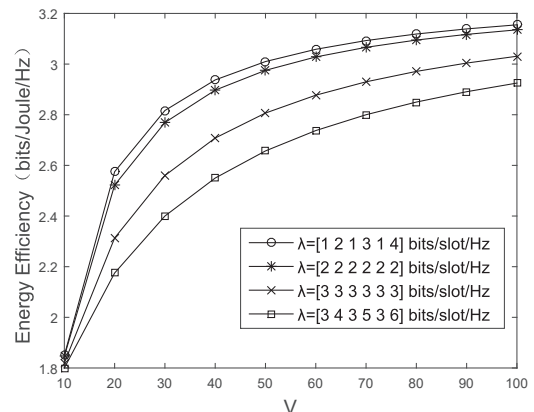


Fig. 3. Energy efficiency versus V and λ .

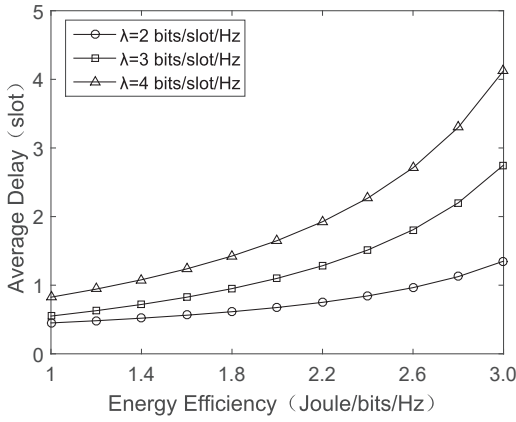


Fig. 4. Average delay versus Energy efficiency.

Furthermore, we can easily get the following result when $D_k(t) \geq \beta$, or

$$\bar{Q} = \sum_{k=1}^K \gamma_k \beta. \quad (42)$$

Combining both (41) and (42) completes the proof. ■

Remarks on EE-delay tradeoff: The bounds in (27) show that U_{EE} increases with V and approaches to U_{EE}^{opt} as $V \rightarrow \infty$. (37) indicates that the time average queue backlog bound increases linearly with V within the limit of β , which can also reveal the relationship between delay and V by Little's Theorem [13]. Furthermore, from these two theorems we can see that a higher EE can be achieved at the cost of a longer delay, and there is clearly a tradeoff between the EE and delay.

V. SIMULATION RESULTS AND DISCUSSIONS

Let us consider a scenario with six virtual BSs and six terminals accessing BSs randomly with different parameters. We set $P^C = [P_1^C, P_2^C, P_3^C, P_4^C, P_5^C, P_6^C] = [0.2, 0.3, 0.2, 0.4, 0.3, 0.2]$ Watt for virtual BS's 1~6 and $R^{av} = [R_1^{av}, R_2^{av}, R_3^{av}, R_4^{av}, R_5^{av}, R_6^{av}] = [2, 3, 1, 4, 2, 1]$ bits/slot/Hz for terminals 1~6. Moreover, we set $P_{max} = 20$ W as the maximum total power that all virtual BSs can reach. In addition, it is assumed that the wireless channel gain $S(t)$ obeys a uniform distribution from 5 to 14 with an expression of $S(t)/\rho^2 \sim U(5, 14)$. We use 10,000 slots to approximate $t \rightarrow \infty$.

Fig. 2 illustrates the impact of data arrival rate λ on average delay with the setting of $\beta = 50$ and V varying from 10 to 90. When λ is less than seven, the average delay increases exponentially with respect to λ . When $\lambda \geq 7$, the average delay tends to be saturated because the network queueing length goes beyond the limit of equivalent queue length, and new arriving data will be refused to enter the system. Therefore, queue lengths will converge; so will the delay. In addition, it is easy to see that the increase of V has a negative impact on the average delay.

Fig. 3 shows the EE of the system versus the control parameter V with different settings of λ . The first two curves exhibit almost the same trend because of the same total network average data arrival rate $\lambda = 12$, and their EE's

are higher than the others, which reveals that the EE of a WVN is not affected by the differences in user requirements, and related only to the total traffic amount of network, that is, the WVN can achieve a traffic balance between different virtual BSs due to its global dynamical resource allocation. Furthermore, λ has a negative impact on U_{EE} .

Fig. 4 shows the average delay of the system versus EE with different settings of λ . The increase of EE will lead to a moderate increase of the average delay when the data arrival rate is low; while a substantial delay is observed with a high data arrival rate. Moreover, the increase of λ exerts a negative effect on the average delay as well as EE.

VI. CONCLUSION

In this paper, we investigated an EE-delay tradeoff problem in wireless networks virtualization. We formulated the problem as a stochastic optimization problem to maximize its EE with an average delay constraint. We solved this problem using Lyapunov optimization and fractional programming methods, and derived the theoretical bounds for both EE and delay. Specifically, we demonstrated that a performance tradeoff can be achieved through adjusting the corresponding parameters properly.

ACKNOWLEDGMENT

This work was supported in part by National Natural Science Foundation of China (61372070), Intergovernmental International Cooperation on Science and Technology Innovation (2016YFE0123200), and the 111 Project (B08038). The work of G. Zheng was supported by the UK EPSRC under grant number EP/N007840/1.

REFERENCES

- [1] C. Liang and F. R. Yu, "Wireless Network Virtualization: A Survey, Some Research Issues and Challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 358-380, First quarter 2015.
- [2] C. She and C. Yang, "Energy Efficiency and Delay in Wireless Systems: Is Their Relation Always a Tradeoff?," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7215-7228, Nov. 2016.
- [3] C. Xiong, G. Y. Li, Y. Liu, Y. Chen, and S. Xu, "Energy-Efficient Design for Downlink OFDMA with Delay-Sensitive Traffic," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 3085-3095, June 2013.
- [4] L. Musavian and T. Le-Ngoc, "Energy-Efficient Power Allocation Over Nakagami Fading Channels Under Delay-Outage Constraints," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4081-4091, Aug. 2014.
- [5] L. Zhang, J. Liu, and K. Yang, "Quality of Service Modeling of Virtualized Wireless Networks: A Network Calculus Approach," *Springer Mobile Networks and Applications*, vol. 19, no. 4, pp. 572-582, 2014.
- [6] K. Wang, H. Li, F. R. Yu, and W. Wei, "Virtual Resource Allocation in Software-Defined Information-Centric Cellular Networks with Device-to-Device Communications and Imperfect CSI," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10011-10021, Dec. 2016.
- [7] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*. San Rafael, CA, USA: Morgan and Claypool, 2010.
- [8] A. Zappone, E. A. Jorswieck, *Energy efficiency in wireless networks via fractional programming theory*. Now Publishers, Foundations and Trends in Communications and Information Theory, 2015.
- [9] D. Xue and E. Ekici, "Power optimal control in multihop wireless networks with finite buffers," *IEEE Trans. Veh. Technol.*, vol. 62, no. 3, pp. 1329-1339, 2013.
- [10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

- [11] K. Liang, L. Zhao, K. Yang and X. Chu, "Online Power and Time Allocation in MIMO Uplink Transmissions Powered by RF Wireless Energy Transfer," *IEEE Trans. Veh. Technol.*, vol. no. 99, pp. 1-1, 2017.
- [12] M. J. Neely, "Dynamic Optimization and Learning for Renewal Systems," *IEEE Trans. Autom. Control.*, vol. 58, no. 1, pp. 32-46, Jan. 2013.
- [13] Dimitri P. Bertsekas and Robert G. Gallager, *Data Networks (2nd edition)*. Prentice Hall, 1992.
- [14] P. Giaccone, E. Leonardi, and D. Shah, "Throughput region of finite-buffered networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 18, no. 2, pp. 251-263, 2007.
- [15] L. B. Le, E. Modiano, and N. Shroff, "Optimal control of wireless networks with finite buffers," *IEEE/ACM Trans. Netw.*, vol. 20, no. 4, pp. 1316-1329, 2012.
- [16] D. Xue and E. Ekici, "Power optimal control in multihop wireless networks with finite buffers," *IEEE Trans. Veh. Technol.*, vol. 62, no. 3, pp. 1329-1339, 2013.