

Research Article

Congestion-Optimal WiFi Offloading with User Mobility Management in Smart Communications

Bin Liu ¹, Qi Zhu ¹, Weiqiang Tan,² and Hongbo Zhu ¹

¹Jiangsu Key Laboratory of Wireless Communications, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

²School of Computer Science and Educational Software, Guangzhou University, Guangzhou 510006, China

Correspondence should be addressed to Hongbo Zhu; zhuhb@njupt.edu.cn

Received 12 April 2018; Accepted 11 June 2018; Published 1 August 2018

Academic Editor: Mu Zhou

Copyright © 2018 Bin Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We study the WiFi offloading problem in smart communications and adaptively seek for the optimal offloading strategies with the consideration of the mobility management and the dynamical nature of network state. With users mobility management, we formulate the offloading ratio optimization problem based on Markov process. Then, we propose a novel Congestion-Optimal WiFi Offloading (COWO) algorithm based on subgradient method, which aims to obtain the optimal offloading ratio for each access point (AP) to maximize the throughput and minimize the network congestion. Due to the computational complexity of subgradient method, we further improve the COWO algorithm by the equivalent transformation. By viewing all the APs as one virtual WiFi network, we try to optimize the identical offloading ratio for virtual WiFi network and develop a Virtualized Congestion-Optimal WiFi Offloading (VCOWO) algorithm with lower complexity. Under the equivalent conditions, the performance of the VCOWO algorithm could well approximate the optimal results obtained by the COWO algorithm. It is found that the VCOWO algorithm could obtain the upper bound of multiple APs WiFi offloading performance. Moreover, we investigate the impacts of user mobility on the WiFi offloading performance. Simulation results show that the proposed algorithm could achieve higher throughput with lower network congestion compared with other current offloading schemes.

1. Introduction

With the proliferation of smart devices such as smartphones and tablets, cellular networks are facing an exponential growth of mobile data traffic. According to Cisco's forecast, global mobile data traffic is expected to grow to 49 megabytes per month by 2021, a sevenfold increase over 2016 [1–4]. With the limited licensed bandwidth, the cellular network capacity, however, can not keep up with the explosive data growth [5–9]. The mobile operators have been seeking for the cost-effective and timely solution to alleviate the cellular network. Thanks to the abundant unlicensed spectrum and large-scale WiFi deployment, Wireless Local Area Network (WLAN) has attracted much attention as a promising approach to offload data from the cellular network [10–13] and enhance network survivability and resilience in smart communications.

Previous works have demonstrated WiFi offloading prospects in leveraging traffic load [14–18]. The work in [18] proposed the on-the-spot offloading (OTSO) scheme

and showed that the OTSO offloading could leverage more than 65% traffic from the cellular network through an experiment in Seoul. Gass and Diot [19] further verified that the WiFi offloading is favorable even if the connecting time is insufficient. Much effort has also been seen in WiFi offloading schemes and performance analysis under the integrated cellular and WLAN networks framework [20]. The authors in [21] demonstrated that larger portions of cellular traffic could be offloaded to WiFi if the delay to wait for WiFi network is allowed during user movement. It presented the offloading scheme called Wiffler, which schedules the network access based on the historical access knowledge. By extracting typical users' mobility profiles, the work of [22] optimized the energy consumption in offloading. The authors in [23] studied the capacity of delayed offloading without prior knowledge of users' mobility patterns and proposed online scheduling policy to maximize the amount of offloaded data. By taking downloading cost and delay into the offloading decision, the delayed offloading scheme was

proposed in [24] to harvest maximum benefits from WiFi offloading. The optimal transmission deadline for delayed offloading was further derived in [25] and could achieve the maximum monetary incentive while maintaining the outage probability. The WiFi offloading and LTE WLAN Aggregation (LWA) was jointly considered in [26] to make the full use of the spectrum in the licensed and unlicensed carriers aggregation approach.

However, previously proposed schemes mainly focus on using WiFi offloading merely as a capacity-augment solution and try to offload as much data traffic to WiFi as possible, without systematically considering the dynamic nature of the network, which may result in the network congestion and degrade user's quality of experience (QoE). For instance, the OTSO scheme, which has been used as a default setting in most of the smartphones, enables users to handover and offload data through WiFi whenever users enter the network coverage. Rather than consider the load balance and network congestion in the WiFi network, this scheme simply decides the offloading by the availability of the network. It nevertheless increases the access conflicts and degrades the performance [27]. The network selection game was proposed in [28, 29] to analyze the offloading by jointly considering the network congestion and the cost of switching between different networks. The Pareto-efficiency of the equilibria in congestion game was proved in [30], and a client-centric network selection is further proposed to reduce the network congestion. Though the load balance and network congestion are considered, few of these works are involved in the optimal users offloading ratio during their movement or analyze the impacts of network congestion and user mobility on WiFi offloading.

In this paper, we investigate WiFi offloading with consideration of the network congestion and the user mobility management in smart communications and deduce the throughput and blocking probability based on Markov process. From the geometric deduction, the user flow equilibrium is observed, i.e., the rate of the users entering and leaving out of the network is equal, and thus we obtain the user flow rate between different networks. We optimize the ratio of users performing offloading when they move into the WiFi access points (APs), which aims to maximize the network throughput and minimize the network congestion. In this paper, the network congestion is characterized by the blocking probability. To obtain the optimal offloading ratio for each WiFi network, the Congestion-Optimal WiFi Offloading (COWO) algorithm is proposed based on the subgradient method. High computational complexity often occurs, especially in the dense APs deployed scenario. To this end, we make the nearly equivalent transformation and view all the WiFi networks as a virtual network. In this case, the Virtualized Congestion-Optimal WiFi Offloading (VCOWO) algorithm could well approximate to the optimal results under equivalent conditions. Moreover, the effects of users mobility on offloading effectiveness are demonstrated.

The rest of this paper is organized as follows. In Section 2, we describe the scenario of WiFi offloading in Section 2. We analyze the WiFi offloading performance and formulate the offloading optimization problem in Section 3. The

TABLE I: Notations and definitions.

Notation	Definition
S_{tot}	Coverage of the cellular network
S_i	Coverage of the i th AP
M	The number of APs
D	Radius of cellular network coverage
d_i	Radius of i th AP coverage
λ	Total active users arriving rate
λ_c	Active users arriving rate in the cellular network
λ_i	Active users arriving rate in the i th AP
γ_c	Average user flow rate in the cellular network
γ_i	Average user flow rate in the i th AP
μ_c	Average service completing rate in the cellular network
μ_i	Average service completing rate in the i th AP
ω_c	Service arriving rate in the cellular network
ω_i	Service arriving rate in the i th AP
l_c	Service leaving rate in the cellular network
l_i	Service leaving rate in the i th AP
N_c	Average number of users in the cellular network
N_i	Average number of users in the i th AP
P_c^b	Blocking probability in the cellular network
P_i^b	Blocking probability in the i th AP
T_c	Throughput in the cellular network
T_i	Throughput in the i th AP
C_u	Total number of channels in the cellular network
C_i	Total number of channels in the i th AP
v	Average velocity of user mobility
p_i	Offloading probability in i th AP

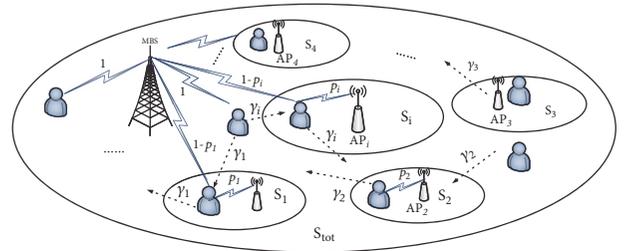


FIGURE 1: WiFi offloading scenario.

Congestion-Optimal WiFi Offloading algorithm is proposed in Section 4. With equivalent conditions, we develop the Virtualized Congestion-Optimal WiFi offloading algorithm in Section 5. We present the numerical results in Section 6 and draw conclusions in Section 7. The symbols used throughout the paper are summarized in Table 1.

2. System Model

We describe the WiFi offloading scenario under the integrated cellular and WLAN networks framework [31–34] in Figure 1, where the MO can tightly integrate the WiFi networks with the cellular networks through the recent IEEE and 3rd Generation Partnership Project (3GPP) standards.

For example, the network discovery and selection functionality (ANDSF) reports the network information related to the access network type, roaming consortium, and venue information through management frames to the macro base station (MBS) [28]. Then, the MBS decides and schedules the user offloading based on the reported information. In this paper, we consider that the MBS covers the whole scenario, with M WiFi APs randomly distributed in its coverage without overlapping with each other. The active users, defined as the users need network service at this moment, are assumed to have arriving rate λ with Poisson arrivals, and the active users follow the uniform distribution on the geometry coverage; i.e., the active user arriving rate for each network is proportional to their coverage, and the active user arriving rates for i th AP and cellular networks are λ_i , $i = 1, 2, \dots, M$, and λ_c , respectively,

$$\begin{aligned} \lambda_i &= \lambda \frac{S_i}{S_{tot}}, \\ \lambda_c &= \lambda \frac{S_{tot} - \sum_{i=1}^M S_i}{S_{tot}}, \end{aligned} \quad (1)$$

where S_{tot} and S_i are the coverage area of the scenario and i th AP, respectively.

The user mobility management is considered, and the offloading ratio in user mobility is further investigated in this paper. With circular coverage, the user flow rates under fluid flow model and geometric angle mobility model were separately deduced in [35–38], and it is proved that the rate of the users entering and leaving the network is equal, which is called user flow equilibrium. With user mobility probability density function $f_V(v)$, we have the user flow rate as

$$\gamma = 2\rho d \int_{-\infty}^{+\infty} v f_V(v) dv = \frac{2NE(v)}{\pi d}, \quad (2)$$

where N is the number of users in the network, d is the radius of the network coverage, and $E(v)$ denotes the first moment of v . We start the analysis of user flow rate with the assumptions as follows:

- (1) All the networks have circular geometric coverage.
- (2) User moves within the coverage area with the random direction; i.e., the angle of direction is uniformly distributed in $[0, 2\pi]$.
- (3) The user's mobility is in uniform linear motion; i.e., $E(v) = v$.

The user flow rate can be written as follows [37]:

$$\gamma = \frac{2NE(v)}{\pi d} = \frac{2Nv}{\pi d}. \quad (3)$$

To be specific, for the i th AP with radius d_i and MBS with radius D , we have the user flow rate separately:

$$\begin{aligned} \gamma_i &= \frac{2N_i v}{\pi d_i}, \\ \gamma_c &= \frac{2N_c v}{\pi D}, \end{aligned} \quad (4)$$

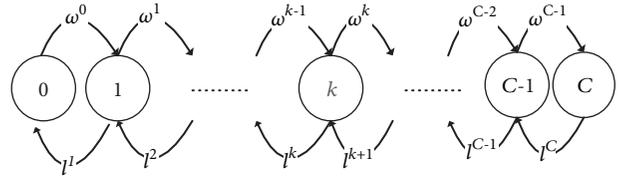


FIGURE 2: Markov state transition.

where N_i and N_c are the numbers of active user in i th AP and MBS.

We formulate the WiFi offloading model based on the discrete Markov process [39–41]. The occupation of network channel is described as a discrete Markov state, the service arriving means that the user enters the network service sequence, and service leaving means that the user cuts off the network service. Each user could access one channel at the same time. The maximum number of users served by the network is constrained by the number of channels C . Let ω^k denote the state transmission rate from the k th state to the $(k+1)$ th state $0 \leq k \leq C-1$, and l^u is the state transition rate of u th state to the $(u-1)$ th state, $1 \leq u \leq C$, as shown in Figure 2. And we have the steady distribution as

$$\begin{aligned} \omega^{k-1} \pi^{k-1} + l^{k+1} \pi^{k+1} &= (\omega^k + l^k) \pi^k, \quad 0 < k < C \\ \mu^{k+1} \pi^{k+1} &= \omega^k \pi^k, \quad k = 0 \\ \omega^{k-1} \pi^{k-1} &= \mu^k \pi^k, \quad k = C \end{aligned} \quad (5)$$

$$\sum_{k=0}^C \pi^k = 1,$$

where π^k is the steady-state probability distribution for the k th state, and

$$\pi^k = \frac{\left(\prod_{m=0}^{k-1} \omega^m / \prod_{x=1}^k l^x \right)}{\left(1 + \sum_{m=1}^C \left(\prod_{n=0}^{m-1} \omega^n / \prod_{x=1}^m l^x \right) \right)}. \quad (6)$$

We try to find the optimal offloading ratio p_i , aiming to maximize the network throughput and minimize the network congestion at both networks when the users enter into the i th AP's coverage area. With offloading ratio p_i , the serve arriving rates at i th AP ω_i and MBS ω_c are defined as

$$\begin{aligned} \omega_i &= \lambda_i + p_i \gamma_i, \\ \omega_c &= \lambda_c + \gamma_c, \end{aligned} \quad (7)$$

where the first term denotes the newly active users in their own coverage, and the second term denotes the service increase by the users flow in their mobility. Note that only p_i of the users will be offloaded to the WiFi network when entering, while all the users moving out of the AP's coverage will be immediately handed over to the cellular network. Assume the average service completing rate in WiFi and

cellular network is μ_i and μ_c , respectively; the service leaving rates for i th AP l_i and MBS l_c are

$$\begin{aligned} l_i &= N_i \mu_i + \gamma_i, \\ l_c &= N_c \mu_c + \sum_{i=1}^M p_i \gamma_i, \end{aligned} \quad (8)$$

where the first part is by the user finishing their transmission and turning to inactive, and the second part is by the user moving out of the network coverage, and handover to another network. Only p_i of the users could be accepted by the i th AP, i.e., only p_i of the users flowing out of the cellular network could leave the MBS service sequence, and the rest of users are still kept in service in the cellular network.

In this paper, we try to optimize the offloading ratio for multiple APs, which aims to increase the throughput and reduce the network congestion. On the one hand, it is an incentive to offload more user to leverage the cellular traffic load and to increase the throughput. On the other hand, the increase of offloading will nevertheless result in the conflicts and degrade the performance. Therefore, the offloading ratio should be carefully designed with the consideration of the network congestion and the user mobility.

3. Problem Formulation

In this section, we analyze the WiFi offloading performance from both cellular and WiFi network sides based on the Markov process and model the network congestion, which is characterized by the blocking probability. Then, we define the system utility as the function of throughput and blocking probability and formulate the offloading ratio optimization problem.

3.1. WiFi Network. First, we analyze the i th AP network with C_i channels in total. From (8), the steady-state probability distribution of i th AP is

$$\pi_i^j = \frac{\left(\sum_{m=0}^{C_i} ((\xi_i)^m / m!)\right)^{-1} (\xi_i)^j}{j!}, \quad 0 \leq j \leq C_i, \quad (9)$$

where

$$\xi_i = \frac{\omega_i}{l_i / N_i} = \frac{\lambda_i + p_i \gamma_i}{\mu_i + \gamma_i / N_i} = \frac{\lambda_i + p_i \gamma_i}{\mu_i + 2\nu / (\pi d_i)}. \quad (10)$$

Thus, we have the number of users served by the i th AP N_i as the average queue length, and

$$\begin{aligned} N_i &= N_i(\xi_i, C_i) = E(k | C_i) = \sum_{k=0}^{C_i} k \cdot \pi_i^k \\ &= \xi_i \cdot \left\{ 1 - \frac{\left(\sum_{k=0}^{C_i} ((\xi_i)^k / k!)\right)^{-1} (\xi_i)^{C_i}}{C_i!} \right\}. \end{aligned} \quad (11)$$

In [42], the blocking probability P_i^b and nonblocking probability P_i^{ub} for the i th AP are defined as

$$P_i^b = 1 - P_i^{ub} = \frac{\left(\sum_{m=0}^{C_i} ((\xi_i)^m / m!)\right)^{-1} (\xi_i)^{C_i}}{C_i!} = P(\xi_i), \quad (12)$$

where $P(\xi_i)$ denotes the fact that the blocking probability depends on the ratio of service arriving and leaving rate ξ_i in (10). Then, the average user number N_i can be rewritten as

$$N_i = N_i(\xi_i, C_i) = \xi_i (1 - P_i^b) = \xi_i P_i^{ub}. \quad (13)$$

With the average user number N_i in (13), the user flow rate can be expressed as

$$\begin{aligned} \gamma_i &= \frac{2N_i(\xi_i, C_i)\nu}{\pi d_i} = \frac{2\xi_i \nu P_i^{ub}}{\pi d_i} \\ &= \frac{2\nu P_i^{ub}}{\pi d_i} \cdot \frac{\lambda_i + p_i \gamma_i}{\mu_i + 2\nu / (\pi d_i)}. \end{aligned} \quad (14)$$

After modification, it is observed that the user flow rate γ_i is the function of blocking probability, which is denoted as $\phi(P_i^b)$, and we have

$$\gamma_i = \frac{2\nu \lambda_i P_i^{ub}}{\pi d_i \mu_i + 2\nu (1 - P_i^b)} = \phi(P_i^b). \quad (15)$$

With the user flow rate in (15), we have the blocking probability P_i^b in its implicit function form $P_i^b = \varphi(P_i^b)$ as

$$\begin{aligned} P_i^b &= P(\xi_i) \Big|_{\xi_i = (\lambda_i + p_i \gamma_i) / (\mu_i + 2\nu / (\pi d_i))} \\ &= \frac{\left((\lambda_i + p_i \phi(P_i^b)) / (\mu_i + 2\nu / (\pi d_i))\right)^{C_i} / C_i!}{\sum_{k=0}^{C_i} \left(\left((\lambda_i + p_i \phi(P_i^b)) / (\mu_i + 2\nu / (\pi d_i))\right)^k / k!\right)}, \end{aligned} \quad (16)$$

where the second equation is denoted as $\varphi(P_i^b)$. Clearly, P_i^b is the root of $\psi(x)$ that holds

$$\begin{aligned} \psi(x) &= x - \varphi(x) = x - P(\xi_i) \Big|_{\xi_i = (\lambda_i + p_i \phi(x)) / (\mu_i + 2\nu / (\pi d_i))} \\ &= 0, \end{aligned} \quad (17)$$

and it could be solved by the bisection method.

Proof. See Appendix A. \square

From (16), it is also found that the P_i^b depends on the offloading ratio p_i , i.e., $P_i^b = h_i(p_i)$, and we thus have the user flow rate in the form of a function of p_i

$$\gamma_i = \frac{2\nu \lambda_i (1 - h(p_i))}{\pi d_i \mu_i + 2\nu (1 - p_i + p_i h(p_i))} = \gamma_i(p_i). \quad (18)$$

Then, the throughput achieved under the offloading probability p_i in the i th AP is for simplicity; we normalize the achievable data rate of per channel in the i th AP network as R_i , and the throughput can be expressed as

$$\begin{aligned} T_i(p_i) &= R_i N_i(\xi_i, C_i) \\ &= R_i (1 - h_i(p_i)) \frac{\lambda_i + p_i \gamma_i(p_i)}{\mu_i + 2\nu / (\pi d_i)}. \end{aligned} \quad (19)$$

3.2. *Cellular Network.* The analysis of the cellular network follows the similar approaches in the WiFi network. Consider the cellular network with total C_u channels, and the steady-state probability distribution of the cellular network is expressed as

$$\pi_c^u = \frac{(\xi_c)^u}{\left(\sum_{m=0}^{C_u} ((\xi_c)^m / m!)\right) u!}, \quad 0 \leq u \leq C_u, \quad (20)$$

where

$$\xi_c = \frac{\lambda_c + \gamma_c}{\mu_c + \sum_{i=1}^M (p_i \gamma_i / N_c)}. \quad (21)$$

The length of service queue is equal to the average number of users N_c served by the cellular network,

$$\begin{aligned} N_c(\xi_c) &= E(u) = \sum_{u=0}^{C_u} u \cdot \pi_c^u \\ &= \xi_c \left\{ 1 - \frac{\left(\sum_{m=0}^{C_u} ((\xi_c)^m / m!)\right)^{-1} (\xi_c)^{C_u}}{C_u!} \right\} \\ &= \xi_c (1 - P_c^b) = \xi_c P_c^{ub}, \end{aligned} \quad (22)$$

where P_c^b and P_c^{ub} are defined the blocking probability and nonblocking probability in the cellular network as (13):

$$P_c^b = 1 - P_c^{ub} = \frac{(\xi_c)^{C_u}}{\left(\sum_{m=0}^{C_u} ((\xi_c)^m / m!)\right) C_u!} = P_c(\xi_c), \quad (23)$$

where $P(\xi_c)$ denotes that the blocking probability depending on the ratio of service arriving and leaving rate ξ_c .

Fixing the average user number N_c in (22), the user flow rate γ_c can be expressed as the function of blocking probability, which is denoted as $\phi_c(P_c^b)$,

$$\gamma_c = \frac{N_c(\xi_c) \nu}{\pi D} = \frac{2\nu (P_c^{ub} \lambda_c - \sum_{i=1}^M p_i \gamma_i)}{(\pi D \mu_c - 2P_c^{ub} \nu)} = \phi_c(P_c^b). \quad (24)$$

With the user flow rate in (24), we have the blocking probability P_c^b in its implicit function form,

$$P_c^b = P_c(\xi_c) \Big|_{\xi_c = (\lambda_c + \phi_c(P_c^b)) / (\mu_c + \sum_{i=1}^M (p_i \gamma_i / N_c))} = \varphi_c(P_c^b). \quad (25)$$

Similar to (16), P_c^b could be obtained through solving (25) by the bisection method. From (25), it is also found that the P_c^b depends on the offloading ratio for each AP p_i , $i = 1, 2, M$; i.e., P_c^b is the function of vector composed by p_i , $P_c^b = g(\vec{p})$, $\vec{p} = [p_1, p_2, \dots, p_i, \dots, p_M]$. Thus, the user flow rate γ_c also has the formation of the function of offloading ratio vector \vec{p} as

$$\gamma_c = \frac{2\nu \left\{ [1 - g(\vec{p})] \lambda_c - \sum_{i=1}^M p_i \gamma_i \right\}}{[\pi D \mu_c - 2 [1 - g(\vec{p})] \nu]} = \gamma_c(\vec{p}). \quad (26)$$

The throughput of the cellular network is similarly given:

$$T_c(\vec{p}) = R_c (1 - g(\vec{p})) \cdot \frac{\pi D \left\{ (1 - g(\vec{p})) \lambda_c - \sum_{i=1}^M p_i \gamma_i \right\}}{(1 - g(\vec{p})) \cdot [\pi D \mu_c - 2\nu (1 - g(\vec{p}))]}, \quad (27)$$

where R_c is the normalized channel rate of the cellular network.

3.3. *Congestion Optimization Problem.* In this paper, we seek for the optimal offloading ratio for multiple APs in user mobility, which aims to maximize the throughput and minimize the network congestion. The throughput is the sum of all networks, i.e., $T(\vec{p}) = \sum_{i=1}^M T_i(p_i) + T_c(\vec{p})$, and the network congestion is characterized by the blocking probability, i.e., $B(\vec{p}) = \sum_{i=1}^M P_i^b + P_c^b$. Thus, we set the system utility function $Q(\vec{p})$ so that it increases with the throughput and decreases with the network congestion, where the blocking probability acts as a penalty for network congestion. The constant β is defined as weight over the throughput and blocking probability, which embodies the sensitivity to network congestion. The system utility function could be formulated as

$$\begin{aligned} \max_{\vec{p}} \quad & Q(\vec{p}) = T(\vec{p}) - \beta B(\vec{p}) \\ &= \sum_{i=1}^M R_i N_i(\xi_i, C_i) + R_c N_c(\xi_c) \\ &\quad - \beta \left(\sum_{i=1}^M P_i^b + P_c^b \right) \end{aligned} \quad (28)$$

$$\text{s.t.} \quad \vec{p} = [p_1, p_2, \dots, p_i, \dots, p_M],$$

$$0 \leq p_1 \dots p_i \dots p_M \leq 1$$

4. Congestion-Optimal WiFi Offloading

In this section, we proposed the COWO algorithm to obtain the optimal offloading ratio p_i for each AP, which is based on subgradient method [43] and further reveal the impacts of user mobility on WiFi offloading performance.

4.1. *Congestion-Optimal WiFi Offloading Algorithm.* The optimization problem in (28) is convex function with the convex feasible region. Thus, the subgradient method can be used to solve the optimization problem.

Proof. See Appendix B and Appendix C. \square

The augmented Lagrangian function of optimization problem (28) is written as

$$\begin{aligned} \max \quad & L(\vec{p}, \nu) \\ &= \sum_{i=1}^M \left[(1 - P_i^b) \frac{\lambda_i + p_i \gamma_i(p_i)}{\mu_i + 2\nu / (\pi d_i)} - \beta P_i^b \right] \end{aligned}$$

$$\begin{aligned}
& + \frac{\pi D (P_c^{ub} \lambda_c - \sum_{i=1}^M p_i \gamma_i)}{\pi D \mu_c - 2\nu P_c^{ub}} - \beta P_c^b \\
& - \sum_{i=1}^M \nu_i (p_i - 1) \\
\text{s.t. } & p_1 \dots p_i \dots p_M \geq 0, \\
& \nu_i > 0,
\end{aligned} \tag{29}$$

where ν_i is the Lagrange multiplier. The Lagrange problem in (29) could be solved in subgradient method, and the optimal $\vec{\mathbf{p}}$ satisfies

$$\begin{aligned}
& \frac{\partial L(\vec{\mathbf{p}}, \alpha)}{\partial \vec{\mathbf{p}}} \\
& = \left[\frac{\partial L(\vec{\mathbf{p}}, \alpha)}{\partial p_1} \quad \frac{\partial L(\vec{\mathbf{p}}, \alpha)}{\partial p_2} \quad \dots \quad \frac{\partial L(\vec{\mathbf{p}}, \alpha)}{\partial p_M} \right]^T \tag{30} \\
& = \vec{\mathbf{0}},
\end{aligned}$$

where the i th element in the partial derivatives of p_i can be calculated as (31), as shown below.

$$\begin{aligned}
\frac{\partial L(\vec{\mathbf{p}}, \nu)}{\partial p_i} & = \left[R_i (1 - P_i^b (1 + C_i - \xi_i P_i^{ub})) \right. \\
& \quad \left. - \frac{\beta_i P_i^b (C_i - \xi_i P_i^{ub})}{\xi_i} \right] \frac{\partial \xi_i}{\partial p_i} \\
& + \left[R_i (1 - P_c^b (1 + C_u - \xi_c P_c^{ub})) \right. \\
& \quad \left. - \frac{\beta_c P_c^b (C_u - \xi_c P_c^{ub})}{\xi_c} \right] \frac{\partial \xi_c}{\partial p_i} + \nu_i.
\end{aligned} \tag{31}$$

Equation (30) is nonlinear. Thus, we make the transformation to simplify the calculation. The partial derivative of p_i is written as

$$\frac{\partial L(\vec{\mathbf{p}}, \nu)}{\partial p_i} = \gamma_i W^i - \nu_i \tag{32}$$

where

$$\begin{aligned}
W^i & = \frac{R_i}{\gamma_i} \frac{\partial N_i(\xi_i, C_i)}{\partial \xi_i} \frac{\partial \xi_i}{\partial p_i} + \frac{R_i \beta}{\gamma_i} \frac{\partial P_i^{ub}}{\partial \xi_i} \frac{\partial \xi_i}{\partial p_i} \\
& + \frac{R_c}{\gamma_i} \frac{\partial N_c(\xi_c)}{\partial \xi_c} \frac{\partial \xi_c}{\partial p_i} + \frac{R_c \beta}{\gamma_i} \frac{\partial P_c^{ub}}{\partial \xi_c} \frac{\partial \xi_c}{\partial p_i} > 0.
\end{aligned} \tag{33}$$

To achieve the identity $\partial L(\vec{\mathbf{p}}, \nu) / \partial \vec{\mathbf{p}} = \vec{\mathbf{0}}$, $\nu_i = \gamma_i W^i$ for $i = 1, 2, \dots, M$. Thus, we have

$$W^1 = \dots = W^M = \chi, \tag{34}$$

where $\chi = \nu_i / \gamma_i$, and the augmented Lagrangian function in (28) can be written as

$$\begin{aligned}
\max \quad & L(\vec{\mathbf{p}}, \chi) \\
& = \sum_{i=1}^M \left[(1 - P_i^b) \frac{\lambda_i + P_i \gamma_i (p_i)}{\mu_i + 2\nu / (\pi d_i)} - \beta_i P_i^b \right] \\
& + \frac{\pi D (P_c^{ub} \lambda_c - \sum_{i=1}^M p_i \gamma_i)}{\pi D \mu_c - 2\nu P_c^{ub}} - \beta_c P_c^b \tag{35} \\
& - \chi \sum_{i=1}^M \nu_i (p_i - 1) \quad \chi > 0
\end{aligned}$$

$$\text{s.t. } p_1 \dots p_i \dots p_M \geq 0.$$

Then, χ could be updated with pace ϑ , and we have

$$\chi^{(k+1)} = \left[\chi^{(k)} + \vartheta_i^{(k)} \sum_{i=1}^M \nu_i (p_i - 1) \right]^+ \tag{36}$$

The following parts give the update process in detail. For i th AP, the service arriving rate is $\omega_i = \lambda_i + p_i \gamma_i$, and $\lambda_i \leq \omega_i \leq \lambda_i + \gamma_i$. It is observed from (37) that W^i is dependent on ω_i , and thus we define the function $f_i(\omega_i)$

$$\frac{\partial L(\vec{\mathbf{p}}, \chi)}{\partial p_i} = \gamma_i (W^i - \chi) = f_i(\omega_i). \tag{37}$$

It can be proved that the $f_i(\omega_i)$ is a decreasing function of ω_i by the partial derivative deduction. Thus, if the zero point of $f_i(\omega_i)$ exists, it can be obtained by the bisection method. Let us denote the zero point of $f_i(\omega_i)$ as ω_i^* . If $f_i(\omega_i)$ does not have zero points, we modify it as

$$\omega_i^* = \begin{cases} \omega_i, & \exists \omega_i \in [\lambda_i, \lambda_i + \gamma_i] \cap f_i(\omega_i) = 0 \\ \lambda_i, & \forall \omega_i \in [\lambda_i, \lambda_i + \gamma_i] \cap f_i(\omega_i) > 0 \\ \lambda_i + \gamma_i, & \forall \omega_i \in [\lambda_i, \lambda_i + \gamma_i] \cap f_i(\omega_i) < 0. \end{cases} \tag{38}$$

If $\forall \omega_i \in [\lambda_i, \lambda_i + \gamma_i]$, $f_i(\omega_i) > 0$, the offloading ratio p_i is obtained by

$$p_i \gamma_i = \omega_i^* - \lambda_i, \quad i = 1, 2, \dots, M. \tag{39}$$

When $\forall \omega_i \in [\lambda_i, \lambda_i + \gamma_i]$, $f_i(\omega_i) < 0$, the offloading ratio is modified as

$$p_i^* = \begin{cases} p_i, & \exists p_i \in [0, 1] \cup \frac{\partial L(\vec{\mathbf{p}}, \alpha)}{\partial p_i} = 0 \\ 0, & \forall p_i \in [0, 1] \cup \frac{\partial L(\vec{\mathbf{p}}, \alpha)}{\partial p_i} > 0 \\ 1, & \forall p_i \in [0, 1] \cup \frac{\partial L(\vec{\mathbf{p}}, \alpha)}{\partial p_i} < 0. \end{cases} \tag{40}$$

Based on the subgradient method, the implementation of the proposed COWO algorithm is provided as in Algorithm 1.

```

1: Initialize:  $\vec{\mathbf{p}}^{(0)} = [p_1^0, p_2^0 \dots p_M^0] = [0, 0 \dots 0] = \vec{\mathbf{0}}, \vartheta_i^{(0)} = 1, k = 0$ 
2: while  $|\chi^{(k+1)} - \chi^k| \geq \varepsilon$  do
3:   Calculate  $P_i^b$  for  $1 \leq i \leq M$  and  $P_c^b$  under the initial  $\vec{\mathbf{p}}^{(k)}$  in the bisection method
4:   Get initial updating factor  $\chi^{(0)} = [(1/M) \sum_{i=1}^M W^i]^*$ 
5:   Solve equation (37) and get  $\omega_i^*$ 
6:   Update  $\omega_i$  by (38)
7:   Solve out  $\vec{\mathbf{p}}^{(k)}$ 
8:   Set  $k = k + 1$  and update  $\chi^{(k+1)}$  according to (36)
9:   Iterate  $\vec{\mathbf{p}}^{(k)}$  into  $\vec{\mathbf{p}}^{(k+1)}$ 
10: end while
11: Set the optimal offload factor vector  $\vec{\mathbf{p}}^* = \vec{\mathbf{p}}^{(k+1)}$ 

```

ALGORITHM 1: The proposed COWO algorithm.

4.2. Impacts of User Mobility on WiFi Offloading. In this section, we study the impacts of user mobility, mainly characterized by average velocity, on the WiFi offloading. First, we show its impacts on WiFi offloading throughput. The derivatives of throughput T_i with respect to v are

$$\begin{aligned} \frac{\partial T_i}{\partial v} &= R_i \frac{\partial N_i(\xi_i, C_i)}{\partial v} = R_i \frac{\partial \xi_i}{\partial v} \left(1 - P_i^b - \xi_i \frac{\partial P_i^b}{\partial \xi_i} \right) \\ &= R_i \frac{\partial \xi_i}{\partial v} \left(1 - P_i^b (C_i + 1 - \xi_i P_i^{\mu b}) \right). \end{aligned} \quad (41)$$

From (10) and (11), ξ_i can be expressed as

$$\xi_i = \frac{\lambda_i + 2p_i N_i(\xi_i, C_i) v / (\pi d_i)}{\mu + 2v / (\pi d_i)}, \quad (42)$$

and we take the derivatives ξ_i of v , and we have

$$\begin{aligned} \frac{\partial \xi_i}{\partial v} &= 2 \frac{v p_i (\partial N_i(\xi_i, C_i) / \partial v) + p_i N_i(\xi_i, C_i)}{\pi d_i (\mu + 2v / \pi d_i)} \\ &\quad - \frac{(\lambda_i + 2p_i N_i(\xi_i, C_i) v / \pi d_i) (\mu + 2v / \pi d_i)^{-1}}{\pi d_i (\mu + 2v / \pi d_i)}. \end{aligned} \quad (43)$$

Thus, the derivatives of throughput T_i with respect to v could be rewritten as

$$\begin{aligned} \frac{\partial T_i}{\partial v} &= 2R_i \left(1 - (1 - P_i^b (C_i + 1 - \xi_i P_i^{\mu b})) \frac{2v p_i / (\pi d_i)}{\mu + 2v / (\pi d_i)} \right)^{-1} \\ &\quad \cdot \left(\frac{p_i N_i(\xi_i, C_i) - \xi_i}{\pi d_i (\mu + 2v / (\pi d_i))} \right) (1 - P_i^b (C_i + 1 - \xi_i P_i^{\mu b})), \end{aligned} \quad (44)$$

where the details about the computation and the associated analysis can be found in the literature [44–46]. With the fact that $N_i(\xi_i, C_i) = \xi_i P_i^{\mu b} < \xi_i$ and the deduction in Appendix A and B, we have

$$0 < 1 - P_i^b (C_i + 1 - \xi_i P_i^{\mu b}) < 1. \quad (45)$$

Therefore, $\partial T_i / \partial v < 0$, which draws the conclusion that the average throughput for each network T_i is decreasing with the velocity v . It is also observed that ξ_i decreases with v . Furthermore, the blocking probability P_i^b has

$$\frac{\partial P_i^b}{\partial v} = \frac{P_i^b}{\xi_i} (C_i - \xi_i P_i^{\mu b}) \frac{\partial \xi_i}{\partial v} < 0 \quad (46)$$

Thus, P_i^b also decreases with the values of v . Similarly, the similar conclusions can be drawn in the cellular network. The blocking ratio P_c^b and throughput T_c all decrease with velocity. Therefore, the higher speed will degrade the capacity, but could lower the congestion. It could be explained by the fact that the increase of mobility rate in one hand enhances the handover opportunity, as well as the fairness for each user, since they have the more chance to access network; i.e., the increase of user mobility gives more opportunities to enter another network and enables some user who is always failing to connect to the AP to reconsider the network access. It could also lower the blocking probability in the congested situation but can not increase the system capacity.

5. Virtualized Congestion WiFi Offloading Algorithm

The COWO algorithm could obtain the optimal offloading ratio for each AP. However, it is based on the subgradient method, whose computational complexity will drastically increase with the number of networks [47, 48]. In this section, rather than optimize the independent offloading ratio for each AP, we view all the APs as one virtual WiFi network, and the channels and other resources in the WiFi networks will be jointly scheduled. Then, we proposed VCOWO algorithm, which could obtain the approximately optimal results under the equivalent conditions with much lower complexity. Though we did not get the separate offloading ratio, we can obtain some general results and useful insights from it.

Through observation, it is found that the offloading decision mainly depends on the service arriving ω and leaving rate l for given total channels in each network. As all the WiFi networks act as the capacity-augment utilities, it is reasonable

to view all the AP as one virtual WiFi network, where all the channels are centrally controlled. For simplicity, we consider the AP to have the same coverage, i.e., $d_i = d, i = 1, 2, \dots, M$, and same serving completing rate $\mu_i = \mu$. The virtual WiFi network satisfies the equivalent conditions as follows.

- (1) The user flow rate of virtual WiFi network γ_a is the sum of user flow rate for each AP; i.e.,

$$\gamma_a = \sum_{i=1}^M \gamma_i = \sum_{i=1}^M \frac{2N_i v}{\pi d_i} = M \frac{2N_i v}{\pi d} = \frac{2N_i v}{\pi d_{eq}}, \quad (47)$$

where $d_{eq} = d/M$.

- (2) The active users arriving rate of virtual WiFi network λ_a is equal to total active users arriving rate of all APs,

$$\lambda_a = \sum_{i=1}^M \lambda_i = \sum_{i=1}^M \lambda \frac{d_i^2}{D^2} = \lambda M \frac{d^2}{D^2}. \quad (48)$$

- (3) All the channels in each AP are scheduled by the virtual WiFi network, i.e., the virtual WiFi network C_a has $C_a = \sum_{i=1}^M C_i$.

- (4) The user serving rate in virtual WiFi network is $\mu_a = \mu$.

We try to find the optimal offloading ratio p_a for the virtual WiFi network. Thus, the steady-state distribution of virtual WiFi network can be expressed as

$$\pi_a^j = \frac{(\sum_{m=0}^{C_a} ((\xi_a)^m / m!))^{-1} (\xi_a)^j}{j!}, \quad 0 \leq j \leq C_a, \quad (49)$$

where

$$\xi_a = \frac{\lambda_a + p_a \gamma_a}{\mu_a + 2v / (\pi d_{eq})}. \quad (50)$$

Therefore, the number of users served by the virtual WiFi network N_a is the average queue length, and

$$\begin{aligned} N_a(\xi_a) &= \sum_{k=0}^{C_a} k \pi_a^k \\ &= \xi_a \left\{ 1 - \frac{(\xi_a)^{C_a}}{(\sum_{k=0}^{C_a} ((\xi_a)^k / k!)) C_a!} \right\}. \end{aligned} \quad (51)$$

Similarly, the blocking probability P_a^b and unblocking probability P_a^{ub} in virtual WiFi network are defined as

$$P_a^b = 1 - P_a^{ub} = \frac{(\sum_{m=0}^{C_a} ((\xi_a)^m / m!))^{-1} (\xi_a)^{C_a}}{C_a!}. \quad (52)$$

Then the users flow rate in virtual network γ_a is written as

$$\begin{aligned} \gamma_a &= \frac{2N_a(\xi_a)v}{\pi d_{eq}} = \frac{2\xi_a v P_a^{ub}}{\pi d_{eq}} \\ &= \frac{2v P_a^{ub}}{\pi d_{eq}} \cdot \frac{\lambda_a + p_a \gamma_a}{\mu_a + 2v / (\pi d_{eq})}. \end{aligned} \quad (53)$$

After modification, it is observed that the user flow rate γ_a is the function of blocking probability, which is denoted as $\phi(P_a^b)$; i.e.,

$$\gamma_a = \frac{2v \lambda_a P_a^{ub}}{\pi \mu_a d_{eq} + 2v(1 - P_a^{ub})} = \phi(P_a^b). \quad (54)$$

Under the equivalent conditions, the analysis of offloading performance in the cellular network could be simplified. The users flowing rate in the cellular network is transformed as γ_c' ,

$$\begin{aligned} \gamma_c' &= \frac{N_c(\xi_c')v}{\pi D} = \frac{2\xi_c' v P_a^{ub}}{\pi D} \\ &= \frac{2v P_c^{ub}}{\pi D} \cdot \frac{\lambda_c + \gamma_c'}{\mu_c + 2v p_a / (\pi D)}, \end{aligned} \quad (55)$$

where

$$\xi_c' = \frac{\lambda_c + \gamma_c'}{\mu_c + 2v p_a / (\pi D)}. \quad (56)$$

The user flow rate γ_a can also be written as the function of blocking probability,

$$\gamma_c' = \frac{2v \lambda_c P_c^{ub}}{\pi D \mu_c + 2v(p_a - P_c^{ub})} = \phi(P_c^b). \quad (57)$$

Similar to (17), the blocking probability could be solved in bisection method [49, 50]. Therefore, the optimization problem in (28) for all multiple APs could be transformed as follows:

$$\begin{aligned} \max \quad & Q'(p_a) \\ &= R_a N_a(\xi_c) + R_c N_c(\xi_c) - \beta(P_a^b + P_c^b) \\ \text{s.t.} \quad & 0 \leq p_a \leq 1. \end{aligned} \quad (58)$$

From (58), it is observed that the optimization is aimed at finding the optimal scalar p_a to maximize the throughput and minimize the blocking probability. The optimal scalar p_a could be solved by the golden section method, and we proposed the VCOWO algorithm as illustrated in Algorithm 2. The computational complexity in COWO is $\mathcal{O}(U + M)$, while the complexity is reduced to $\mathcal{O}(U + 1)$ in VCOWO algorithm, where U is the number of MBS, and $U = 1$ is assumed in this paper. It means the complexity of COWO algorithm increases with the number of APs M , while that of VCOWO algorithm is independent of M . The VCOWO algorithm could approach the optimal results under the equivalent conditions. To be more specific, the VCOWO could achieve the upper bound of the WiFi offloading performance achieved by COWO algorithm. Under the equivalent conditions, all WiFi networks are assumed to serve the users in a cooperation approach and could achieve higher throughput with lower blocking probability with the centralized control of WiFi channels allocation.

```

1: Calculate parameters in equivalent WiFi network
2: Initialize:  $p_0 = 0, k = 0$ 
3: while  $|Q_{k+1} - Q_k| \geq \varepsilon_2$  do
4:   Calculate steady-state probability distribution  $\pi$  under given  $p_k$ 
5:   Calculate utility function  $Q_k$ 
6:   Update  $p_{k+1}$  according to golden section method
7: end while
8: Set the optimal offload factor vector  $p_a^* = p_{k+1}$ .

```

ALGORITHM 2: The proposed VCOWO.

TABLE 2: Simulation parameters.

Parameters	MBS	AP1	AP2	AP3	AP4
$d_i/D(\text{m})$	1000	200	150	100	100
Channels	30	8	7	6	5
R_i/R_c (Mbps)	9.6	13.6	13.6	13.6	13.6
μ (links/s)			1/120		
v (m/s)			8		

6. Numerical Results

In this section, we evaluate the proposed algorithms with the OTSO scheme [18] in system utility as formulated in (28), blocking probability and throughput. The OTSO algorithm enables offloading user to a WiFi network whenever users enter into its coverage, which means the offloading ratio $p = 1$. Further, we show the approximation of VCOWO to COWO under equivalent conditions. Moreover, we verify the impacts of user mobility on WiFi offloading performance. In the simulation, we consider the scenario that contains one MBS $U = 1$ and multiple APs. Each AP is assumed to have no overlap with the others, and other parameters are shown in Table 2.

6.1. Performance Comparison. In this section, we compare the WiFi offloading performance with the multiple APs between COWO and OTSO under user arriving rate ranging from 0 to 0.5. First, Figure 3 illustrates that the system utility varies with the total active user arriving rate λ . When $\lambda < 0.3$, it is found that the system utility grows with λ , and COWO algorithm has similar performance as the OTSO. It is because when the network is unsaturated and the congestion is small, the system utility increases by the throughput bought by the user arriving. The throughput increases since more users flow to the network and make the use of channels when unsaturated. However, it can be observed from Figure 3 that the utility in OTSO decreases with λ . Intuitively, when traffic load in the WiFi network is heavy, offloading data to WiFi network whenever possible will nevertheless incur higher blocking probability, which may degrade the utility function.

To further demonstrate the network congestion in WiFi offloading, we compare the blocking probability in Figure 4. It is obviously observed that the blocking probability in the proposed algorithm is much lower than the OTSO, especially when λ is larger. Without consideration of network congestion, the blocking probability grows rapidly with the

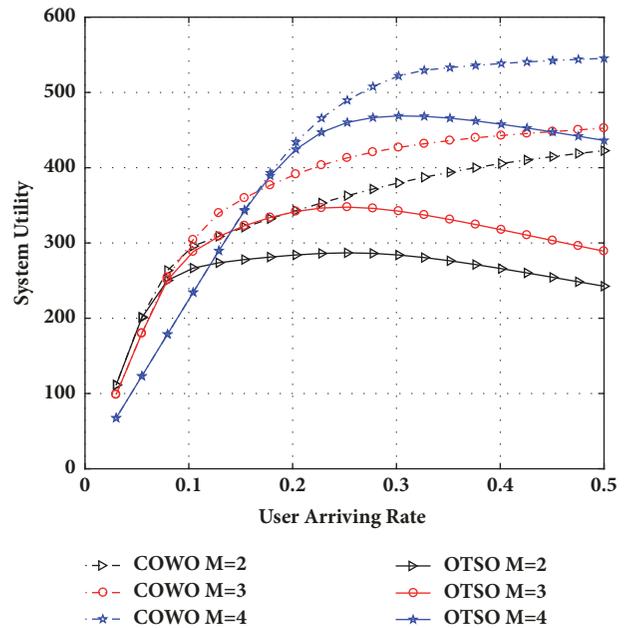


FIGURE 3: System utility varies with the users arriving rate.

active user arriving rate λ , which is also indicated in the decrease of system utility as shown in Figure 3.

Moreover, Figure 5 illustrates that the throughput varies with the user arriving rate λ . It can be observed that the throughput increases with λ . It is because the throughput increases with user arriving when the channels are not fully utilized. When λ is larger, the throughput does not change with the λ . Intuitively, when all the channels have been occupied, the increase of users will not bring higher throughput. Thus, it can be observed from Figure 5, constrained by the number of channels, that the throughput will not increase with the user arriving rate. The throughput is larger with

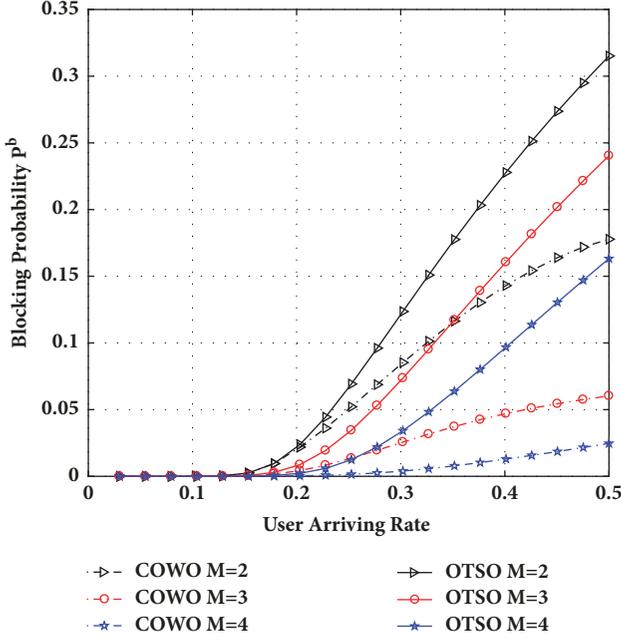


FIGURE 4: Blocking probability varies with the users arriving rate.

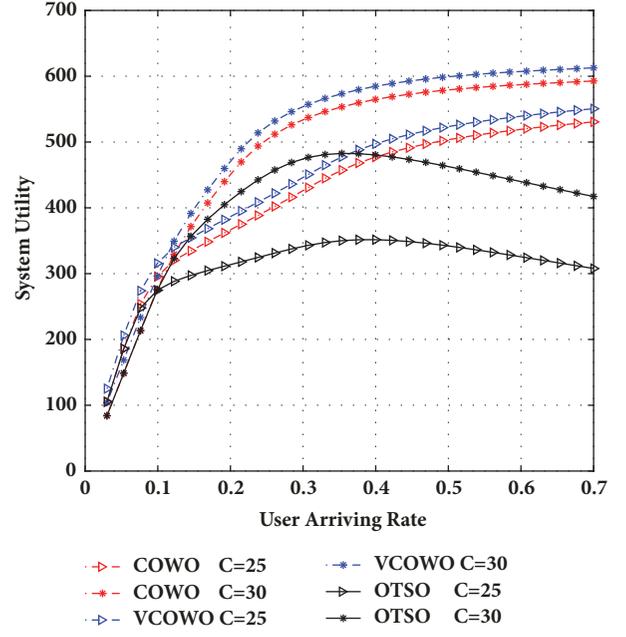


FIGURE 6: System utility comparison varies with the users arriving rate.

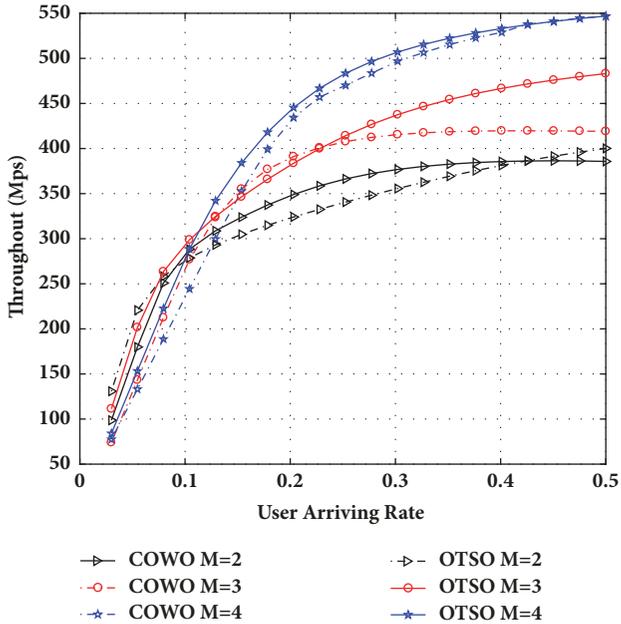


FIGURE 5: Throughput varies with the users arriving rate.

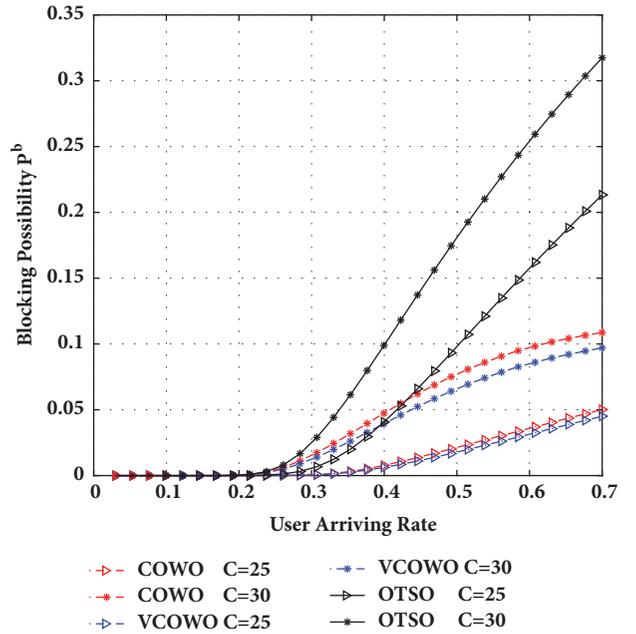


FIGURE 7: Blocking probability comparison varies with the active users arriving rate.

more WiFi APs, e.g., more available channels. To conclude, the proposed algorithm could maximize the throughput with lower blocking probability than OTSO algorithm.

6.2. Approximation Comparison. In this section, we evaluate the VCOWO algorithm under the equivalent conditions. We consider the 4 or 5 APs separately with equal coverage ($d = 150\text{m}$), and each with 5 channels, and it correspondingly has $C = 20$ or 25 channels for the virtual WiFi network. Assume the maximum user arriving rate λ range is 0.7 . Other

parameters all satisfy the equivalent conditions in Section 5. In Figures 6–8, we compare the performance with OTSO under different channel numbers in WiFi network. Clearly, more channels offer service to more users and nevertheless could contribute to the larger system utility. It can be observed that the VCOWO achieves similar system utility, blocking probability and throughput to COWO algorithm. To be more specific, the performance in VCOWO is the

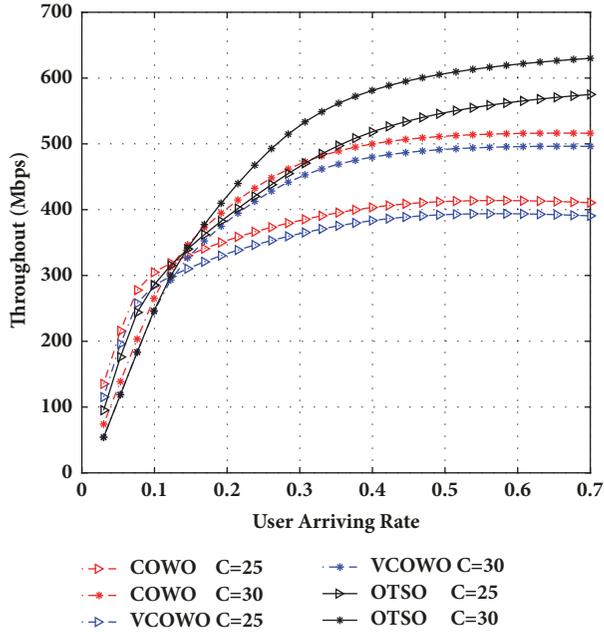
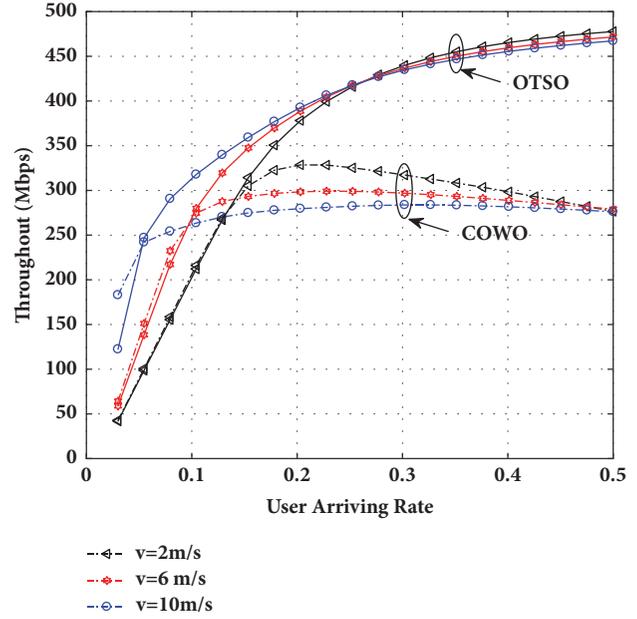
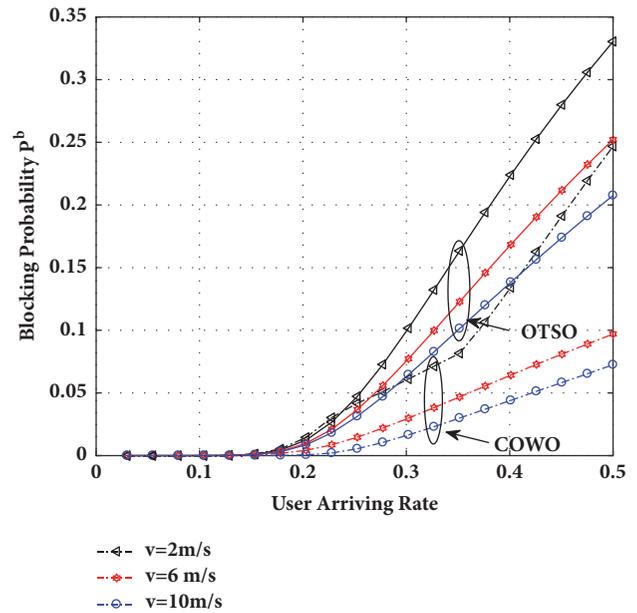


FIGURE 8: Throughput comparison varies with the user arriving rate.

upper bound of COWO under the equivalent situation. Recall the equivalent assumption in Section 5; it is found that all the channels of different APs could be jointly scheduled in the virtual WiFi network. That is to say, all the APs serve the users with the association, rather than separately schedule the users' network access in COWO algorithm. Thus, VCOWO could give the tight upper bound of COWO algorithm performance with much lower complexity under the equivalent conditions.

6.3. User Mobility. In this section, we verify the impacts of user mobility on WiFi offloading as illustrated in Section 5. In Figures 9 and 10, the COWO and OTSO are evaluated under the different velocities with user arriving rate ranging from 0 to 0.5. It can be observed from Figure 9 that the throughput decreases with the velocity. Recall the fact that it was indicated in (41) that throughput T_i is the monotonically decreasing function about the velocity v . Intuitively, we deduce it in another way. At one-moment t , suppose that there are N users in i th AP; the users move with the average velocity v , and the $\xi_i = (\lambda_i + p_i \cdot 2N_i v / (\pi d_i)) / (\mu_i + 2v / (\pi d_i))$. Then, for the next moment, $t + \Delta t$, with the increase of v , we have $\xi_i' = (\lambda_i + p_i \cdot 2N_i (v + \Delta v) / (\pi d_i)) / (\mu_i + 2(v + \Delta v) / (\pi d_i))$, and $\xi_i' > \xi_i$. Therefore, the increase of v reduces the average number of users in the network and thus lowers the throughput and the blocking probability. The user mobility increases the opportunities to access, and could, to some extent, reduce the access blocking in the congested situation. The mobility only gives more users the chances to access the network. However, it can not guarantee the access success or improve the throughput. Actually, constrained by the number of total channels, the capacity will not increase. Even worse, the mobility impedes sustainable connection and nevertheless lowers the throughput in general.


 FIGURE 9: Throughput evaluation under different mobility velocity v .

 FIGURE 10: Blocking probability evaluation under different user mobility velocity v .

7. Conclusion

In this paper, we studied WiFi offloading with user mobility management in smart communications. With consideration of the dynamic network nature, including the network congestion and the user mobility, we proposed the COWO algorithm to optimize the offloading ratio for each AP, which aims to maximize the throughput with lower blocking probability. By viewing all the APs as one virtual WiFi network, the VCOWO algorithm is proposed with

much lower complexity. Under equivalent conditions, the VCOWO algorithm could approximate optimal results by COWO algorithm and give an upper bound of offloading performance of COWO algorithm. Moreover, we investigated the impacts of user mobility on WiFi offloading. It is found that the increase of velocity in user mobility degrades the throughput performance, but could reduce the blocking in some way. For future work, it is an interesting direction for future research to consider other mobility models, especially the heavy-tailed distribution model, which was shown to be more accurate for modeling human mobility. It is promising to deal with the network congestion and load balance in smart communications through artificial intelligence technologies and wireless big data support, which could well predict the user random mobility patterns with statistical information of the possible user's trajectories and schedule the network access in a smarter way.

Appendix

A. Proof of the Bisection Method for Solving P_i^b

Due to P_i^b being the root of $\psi(P_i^b) = 0$, the derivation of $\psi(P_i^b)$ is

$$\psi'(P_i^b) = 1 - \frac{dP(\xi_i)}{d\xi_i} \Big|_{\xi_i=(\lambda_i+p_i\phi(P_i^b))/(\mu+2\nu/(\pi d_i))} \cdot \frac{d\xi_i}{dP_i^b} = 1 \quad (\text{A.1})$$

$$- \frac{\phi'(P_i^b)}{\mu+2\nu/(\pi d_i)} \frac{dP(\xi_i)}{d\xi_i} \Big|_{\xi_i=(\lambda_i+p_i\phi(P_i^b))/(\mu+2\nu/(\pi d_i))}.$$

where

$$\phi'(P_i^b) = \frac{-2\nu[\pi d_i \mu_i + 2\nu(1 - p_i + p_i P_i^b) + p_i]}{[\pi D \mu_i + 2\nu(1 - p_i + p_i P_i^b)]^2} < 0, \quad (\text{A.2})$$

and

$$\frac{dP(\xi_i)}{d\xi_i} = \frac{P_i^b(C_i - \xi_i P_i^{ub})}{\xi_i}. \quad (\text{A.3})$$

The maximum number of users served by the network will not be more than the channel number; i.e., $N(\xi_i) = \sum_{k=0}^{C_i} k \pi_i^k = \xi_i P_i^{ub} \leq C_i \sum_{i=0}^{C_i} \pi^i = C_i$, and thus we have $dP(\xi_i)/d\xi_i \geq 0$. Overall, it is found that $\psi'(P_i^b) \geq 0$, and thus $\psi(P_i^b)$ is an increasing function of P_i^b . Moreover, $P_i^b \in [0, 1]$, $\psi(0) = -\varphi(0) < 0$, and $\psi(1) = 1 - \varphi(1) > 0$. Therefore, $\exists P_i^b \in [0, 1]$, $\psi(P_i^b) = 0$, and the bisection method can be used for the calculation of P_i^b .

B. Proof of Convex Optimization

The utility function can be divided into two parts:

$$\begin{aligned} Q(\vec{\mathbf{p}}) &= T(\vec{\mathbf{p}}) - \beta B(\vec{\mathbf{p}}) \\ &= \sum_{i=1}^M R_i N(\xi_i, C_i) + R_c N_c(\xi_c) \\ &\quad - \beta \left(\sum_{i=1}^N P_i^b + P_c^b \right) \\ &= \sum_{i=1}^M (R_i N(\xi_i, C_i) - \beta P_i^b) \\ &\quad + (R_c N_c(\xi_c) - \beta P_c^b) \\ &= \sum_{i=1}^M Q_i(p_i) + Q_c(\vec{\mathbf{p}}), \end{aligned} \quad (\text{B.1})$$

where the $Q_i(p_i)$ denotes the utility of i th WiFi network and $Q_c(\vec{\mathbf{p}})$ represents the cellular system utility.

The proof of convex optimization in (28) is equal to proving that the utility matrix $Q(\vec{\mathbf{p}})$ is a negative definite matrix. Hessian matrix $\partial^2 Q / \partial \vec{\mathbf{p}}^2$ is

$$\frac{\partial^2 Q}{\partial \vec{\mathbf{p}}^2} = \begin{bmatrix} \frac{\partial^2 Q_1(p_1)}{\partial p_1^2} & 0 & \dots & 0 \\ 0 & \frac{\partial^2 Q_2(p_2)}{\partial p_2^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial^2 Q_M(p_M)}{\partial p_M^2} \\ \frac{\partial^2 Q_c(\vec{\mathbf{p}})}{\partial \vec{\mathbf{p}} \cdot \partial p_1} & \frac{\partial^2 Q_c(\vec{\mathbf{p}})}{\partial \vec{\mathbf{p}} \cdot \partial p_2} & \dots & \frac{\partial^2 Q_c(\vec{\mathbf{p}})}{\partial \vec{\mathbf{p}} \cdot \partial p_M} \end{bmatrix}, \quad (\text{B.2})$$

where

$$\begin{aligned} \frac{\partial^2 Q_c(\vec{\mathbf{p}})}{\partial \vec{\mathbf{p}} \cdot \partial p_1} &= \frac{\partial^2 Q_c(\vec{\mathbf{p}})}{\partial p_1^2} + \frac{\partial^2 Q_c(\vec{\mathbf{p}})}{\partial p_2 \cdot \partial p_1} \dots \\ &\quad + \frac{\partial^2 Q_c(\vec{\mathbf{p}})}{\partial p_M \cdot \partial p_1}. \end{aligned} \quad (\text{B.3})$$

Since $\partial^2 Q_c(\vec{\mathbf{p}})/(\partial p_i \cdot \partial p_j) = \partial^2 Q_c(\vec{\mathbf{p}})/(\partial p_j \cdot \partial p_i), i \neq j$, we have

$$\frac{\partial^2 Q}{\partial \vec{p}^2} = \begin{bmatrix} \frac{\partial^2 Q_1(p_1)}{\partial p_1^2} & 0 & \cdots & 0 \\ 0 & \frac{\partial^2 Q_2(p_2)}{\partial p_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial^2 Q_M(p_M)}{\partial p_M^2} \\ 0 & 0 & \cdots & 0 \end{bmatrix}. \quad (\text{B.4})$$

Then, the proof that the Hessian matrix is negative definite is equal to proving that its eigenvalue is nonnegative. That is,

$$\begin{aligned} \frac{\partial Q_i}{\partial p_i} &= \frac{\partial Q_i}{\partial \xi_i} \frac{\partial \xi_i}{\partial p_i} \geq 0 \\ \frac{\partial^2 Q_i}{\partial p_i^2} &= \frac{\partial^2 Q_i}{\partial \xi_i^2} \left(\frac{\partial \xi_i}{\partial p_i} \right)^2 + \frac{\partial Q_i}{\partial \xi_i} \frac{\partial^2 \xi_i}{\partial p_i^2} \leq 0. \end{aligned} \quad (\text{B.5})$$

We assume that $\partial \xi_i / \partial p_i > 0$ and $\partial^2 \xi_i / \partial p_i^2 < 0$ hold, which has been proved in Appendix B with the bound of P_i^b . Thus, the convex optimization question is converted to prove $\partial Q_i / \partial \xi_i \geq 0$. The first-order and second-order partial derivative of Q_i are, respectively, given by

$$\frac{\partial Q_i}{\partial \xi_i} = 1 - P_i^b (1 + C_i - \xi_i P_i^{\mu b}) - \frac{P_i^b (C_i - \xi_i P_i^{\mu b})}{\xi_i}. \quad (\text{B.6})$$

and

$$\begin{aligned} \frac{\partial^2 Q_i}{\partial \xi_i^2} &= -\frac{P_i^b}{\xi_i} \left[(C_i - \xi_i P_i^{\mu b} + \xi_i P_i^b) (1 + C_i - \xi_i P_i^{\mu b}) - \xi_i \right] \\ &\quad - \frac{P_i^b}{\xi_i^2} \left[(C_i - \xi_i + 2\xi_i P_i^b) (C_i - \xi_i P_i^b) - C_i \right]. \end{aligned} \quad (\text{B.7})$$

The lower and upper bound of P_i^b give the conditions [42]

$$P_i^b \leq \frac{\xi_i}{\xi_i (1 + C_i - \xi_i + P_i^{\mu b}) + C_i} \quad (\text{B.8})$$

$$P_i^b \geq \frac{C_i + \xi_i^2}{\xi_i^2 (1 + C_i - (\xi_i + 1) P_i^{\mu b}) + C_i \xi_i} - \frac{C_i - \xi_i P_i^{\mu b}}{\xi_i}, \quad (\text{B.9})$$

and, thus, we have

$$0 \leq \frac{\partial Q_i}{\partial \xi_i} \leq 1 \quad (\text{B.10})$$

$$\frac{\partial^2 Q_i}{\partial \xi_i^2} \leq 0,$$

which completes the proof.

C. Proof of $\partial^2 \xi_i / \partial p_i^2 < 0$

We prove the $\partial \xi_i / \partial p_i > 0$ and $\partial^2 \xi_i / \partial p_i^2 < 0$ with the bound constraint in Appendix B. From (10), the derivative of ξ_i with p_i is

$$\frac{\partial \xi_i}{\partial p_i} = \frac{\gamma_i}{\mu_i + 2\nu / (\pi d_i)} + \frac{1}{\mu_i + 2\nu / (\pi d_i)} \frac{\partial \gamma_i}{\partial p_i}. \quad (\text{C.1})$$

Then, the first-order derivative of γ_i with p_i is calculated as

$$\begin{aligned} \frac{\partial \gamma_i}{\partial p_i} &= \frac{-2\nu \lambda_i (\pi d_i \mu + 2\nu)}{\pi d_i \mu + 2\nu (1 - p_i P_i^{\mu b})} \frac{\partial P_i^b}{\partial p_i} \\ &\quad + \frac{(2\nu P_i^{\mu b})^2 \lambda_i}{[\pi d_i \mu + 2\nu (1 - p_i P_i^{\mu b})]^2}. \end{aligned} \quad (\text{C.2})$$

With (12), the derivative of P_i^b with p_i is

$$\frac{\partial P_i^b}{\partial p_i} = \frac{P_i^b (C_i - \xi_i P_i^{\mu b})}{\xi_i} \frac{\partial \xi_i}{\partial p_i}. \quad (\text{C.3})$$

With (C.1) and (C.3), the derivative of γ_i with p_i could be rewritten as

$$\begin{aligned} \frac{\partial \xi_i}{\partial p_i} &= \frac{\gamma_i}{\mu_i + 2\nu / (\pi d_i)} \left[1 + \frac{2\nu p_i P_i^{\mu b}}{\pi d_i \mu + 2\nu (1 - p_i P_i^{\mu b})} \right] \\ &\quad \cdot \left[1 + \frac{P_i^b}{\mu_i + 2\nu / (\pi d_i)} \right. \\ &\quad \left. \cdot \frac{2\nu \lambda_i P_i^b (\pi d_i \mu + 2\nu) (C_i - \xi_i P_i^{\mu b})}{[\pi d_i \mu + 2\nu (1 - p_i P_i^{\mu b})] \cdot \xi_i} \right]^{-1}. \end{aligned} \quad (\text{C.4})$$

Since the channel used in $\xi_i P_i^{\mu b}$ is less than the total channel in i -th AP, i.e., $C_i - \xi_i P_i^{\mu b} > 0$, thus, we have $\partial \xi_i / \partial p_i > 0$. Similarly, $\partial^2 \xi_i / \partial p_i^2 < 0$ can be proved.

Data Availability

The authors confirm that the data supporting the findings of this study are available within the article. All the simulation parameters and other experimental details are clearly shown in this article, which could be used to reproduce the experiment related to our findings.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The work of B. Liu, Q. Zhu, and H. Zhu was supported in part by the National Natural Science Foundation of China

(61427801, 61571234, and 61631020) and in part by the Post-graduate Research & Practice Innovation Program of Jiangsu Province (KYCX18.0893). The work of W. Tan was supported in part by the Project of Educational Commission of Guangdong Province of China under Grant 2017KQNCX155.

References

- [1] Cisco, "Cisco Visual Networking Global Mobile Data Traffic Forecast Update, 2016-2021 White Paper".
- [2] B. Liu, Q. Zhu, and H. Zhu, "Delay-Aware LTE WLAN Aggregation for 5G Unlicensed Spectrum Usage," in *Proceedings of the 2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pp. 1-7, Sydney, NSW, June 2017.
- [3] W. Tan, D. Xie, J. Xia, W. Tan, L. Fan, and S. Jin, "Spectral and Energy Efficiency of Massive MIMO for Hybrid Architectures Based on Phase Shifters," *IEEE Access*, vol. 6, pp. 11751-11759, 2018.
- [4] Y. Fang, G. Han, G. Cai, F. C. Lau, P. Chen, and Y. L. Guan, "Design Guidelines of Low-Density Parity-Check Codes for Magnetic Recording Systems," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 1574-1606, 2018.
- [5] F. Shi, W. Tan, J. Xia, D. Xie, L. Fan, and X. Liu, "Hybrid Cache Placement for Physical-Layer Security in Cooperative Networks," *IEEE Access*, vol. 6, pp. 8098-8108, 2018.
- [6] F. Zhou, G. Luo, Y. Liu, Y. Wang, and L. Fan, "Coordinated beamforming for heterogeneous small-cell networks with a non-ideal backhaul," *IET Communications*, vol. 12, no. 5, pp. 595-602, 2018.
- [7] W. Tan, S. Jin, C. Wen, and T. Jiang, "Spectral efficiency of multi-user millimeter wave systems under single path with uniform rectangular arrays," *EURASIP Journal on Wireless Communications and Networking*, vol. 2017, no. 1, 2017.
- [8] L. Fan, X. Lei, N. Yang, T. Q. Duong, and G. K. Karagiannidis, "Secure Multiple Amplify-and-Forward Relaying with Cochannel Interference," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 8, pp. 1494-1505, 2016.
- [9] C. Li, S. Zhang, P. Liu, F. Sun, J. M. Cioffi, and L. Yang, "Overhearing protocol design exploiting intercell interference in cooperative green networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 1, pp. 441-446, 2016.
- [10] Z. Shi, S. Ma, G. Yang, K. Tam, and M. Xia, "Asymptotic outage analysis of HARQ-IR over time-correlated nakagami-m fading channels," *IEEE Transactions on Wireless Communications*, no. 99, article 1, 2017.
- [11] C. Li, Y. Li, K. Song, and L. Yang, "Energy efficient design for multiuser downlink energy and uplink information transfer in 5G," *Science China Information Sciences*, vol. 59, no. 2, pp. 1-8, 2016.
- [12] X. Lai, W. Zou, X. Li, and L. Fan, "Multiuser energy harvesting relaying system with direct links," *IET Communications*, vol. 11, no. 12, pp. 1846-1852, 2017.
- [13] J. Xia, F. Zhou, X. Lai et al., "Cache aided decode-and-forward relaying networks: From the spatial view," *Wireless Communications and Mobile Computing*, vol. 2018, 9 pages, 2018.
- [14] L. Fan, X. Lei, N. Yang, T. Q. Duong, and G. K. Karagiannidis, "Secrecy Cooperative Networks with Outdated Relay Selection over Correlated Fading Channels," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7599-7603, 2017.
- [15] F. Shi, L. Fan, X. Liu, Z. Na, and Y. Liu, "Probabilistic caching placement in the presence of multiple eavesdroppers," *Wireless Communications and Mobile Computing*, vol. 2018, 2018.
- [16] X. Wang, H. Zhang, L. Fan, and Y. Li, "Performance of Distributed Switch-and-Stay Combining for Cognitive Relay Networks with Primary Transceiver," *Wireless Personal Communications*, vol. 97, no. 2, pp. 3031-3042, 2017.
- [17] R. Zhao, Y. Yuan, L. Fan, and Y.-C. He, "Secrecy Performance Analysis of Cognitive Decode-and-Forward Relay Networks in Nakagami-m Fading Channels," *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 549-563, 2017.
- [18] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: how much can wifi deliver?" *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 536-550, 2013.
- [19] R. Gass and C. Diot, "An Experimental Performance Comparison of 3G and Wi-Fi," in *Passive and Active Measurement*, vol. 6032 of *Lecture Notes in Computer Science*, pp. 71-80, Springer, Berlin, Germany, 2010.
- [20] D. Suh, H. Ko, and S. Pack, "Efficiency Analysis of WiFi Offloading Techniques," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3813-3817, 2016.
- [21] B. Aruna, M. Ratul, and A. Venkataramani, "Augmenting mobile 3G using WiFi," in *Pro. Conf. Mobile Syst. Appl. Services*, pp. 209-222, 2010.
- [22] W. Zhang, Y. Wen, and H.-H. Chen, "Toward transcoding as a service: Energy-efficient offloading policy for green mobile cloud," *IEEE Network*, vol. 28, no. 6, pp. 67-73, 2014.
- [23] H. Deng and I.-H. Hou, "On the capacity-performance trade-off of online policy in delayed mobile offloading," *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 526-537, 2017.
- [24] M. H. Cheung and J. Huang, "DAWN: Delay-Aware Wi-Fi offloading and network selection," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1214-1223, 2015.
- [25] H. Ko, J. Lee, and S. Pack, "Performance Optimization of Delayed WiFi Offloading in Heterogeneous Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 9436-9447, 2017.
- [26] B. Liu, Q. Zhu, and H. Zhu, "Delay-Aware LTE WLAN Aggregation in Heterogeneous Wireless Network," *IEEE Access*, vol. 6, pp. 14544-14559, 2018.
- [27] Y. Li, B. Shen, J. Zhang, X. Gan, J. Wang, and X. Wang, "Offloading in HCNs: Congestion-Aware Network Selection and User Incentive Design," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6479-6492, 2017.
- [28] M. H. Cheung, F. Hou, J. Huang, and R. Southwell, "Congestion-aware DNS for integrated cellular and Wi-Fi networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1269-1281, 2017.
- [29] E. Aryafar, A. Keshavarz-Haddad, M. Wang, and M. Chiang, "RAT selection games in HetNets," in *Proceedings of the 32nd IEEE International Conference on Computer Communications (INFOCOM '13)*, pp. 998-1006, Turin, Italy, April 2013.
- [30] A. Keshavarz-Haddad, E. Aryafar, M. Wang, and M. Chiang, "HetNets Selection by Clients: Convergence, Efficiency, and Practicality," *IEEE/ACM Transactions on Networking*, vol. 25, no. 1, pp. 406-419, 2017.
- [31] M. Zhou, Y. Tang, W. Nie, L. Xie, and X. Yang, "GrassMA: graph-based semi-supervised manifold alignment for indoor WLAN localization," *IEEE Sensors Journal*, vol. 17, no. 21, pp. 7086-7095, 2017.
- [32] M. Zhou, Y. Wei, Z. Tian, X. Yang, and L. Li, "Achieving cost-efficient indoor fingerprint localization on wlan platform: a hypothetical test approach," *IEEE Access*, vol. 5, pp. 15865-15874, 2017.

- [33] Y. Fang, G. Han, P. Chen, F. C. M. Lau, G. Chen, and L. Wang, "A Survey on DCSK-Based Communication Systems and Their Application to UWB Scenarios," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1804–1837, 2016.
- [34] J. Li, M. Wen, X. Jiang, and W. Duan, "Space-Time Multiple-Mode Orthogonal Frequency Division Multiplexing with Index Modulation," *IEEE Access*, vol. 5, pp. 23212–23222, 2017.
- [35] S. Pattaramalai, V. A. Aalo, and G. P. Efthymoglou, "Evaluation of call performance in cellular networks with generalized cell dwell time and call-holding time distributions in the presence of channel fading," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 3002–3013, 2009.
- [36] G. Pan, H. Lei, Y. Deng, L. Fan, Y. Chen, and Z. Ding, "On secrecy outage of MISO SWIPT systems in the presence of imperfect CSI," in *Proceedings of the 2016 24th European Signal Processing Conference (EUSIPCO)*, pp. 818–822, Budapest, Hungary, August 2016.
- [37] Z. Shi and Q. Zhu, "Throughput analysis and optimization based on mobility analysis and markov process for heterogeneous wireless networks," *Wireless Personal Communications*, vol. 77, no. 2, pp. 1091–1116, 2014.
- [38] L. Wang and D. Binet, "Mobility-Based Network Selection Scheme in Heterogeneous Wireless Networks," in *Proceedings of the 2009 IEEE 69th Vehicular Technology Conference Spring*, pp. 1–5, Barcelona, Spain, April 2009.
- [39] F. Zhou, L. Fan, X. Lei, G. Luo, H. Zhang, and J. Zhao, "Edge Caching With Transmission Schedule for Multiuser Multirelay Networks," *IEEE Communications Letters*, vol. 22, no. 4, pp. 776–779, 2018.
- [40] X. Lai, W. Zou, D. Xie, and L. Fan, "DF relaying networks in randomly distributed interference environments," in *Proceedings of the 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–6, Nanjing, October 2017.
- [41] F. Zhou, M. Du, Y. Wang, and G. Luo, "Joint source-channel coding for band-limited backhubs in coordinated multi-point systems," *IET Communications*, vol. 10, no. 13, pp. 1562–1570, 2016.
- [42] Z. Shi and Q. Zhu, "Radio resource management scheme for heterogeneous wireless networks based on access proportion optimization," *Journal of Communications and Networks*, vol. 15, no. 5, pp. 527–537, 2013.
- [43] I. Matei and J. S. Baras, "Performance evaluation of the consensus-based distributed subgradient method under random communication topologies," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 754–771, 2011.
- [44] F. Zhou, L. Fan, N. Wang, G. Luo, J. Tang, and W. Chen, "A Cache-Aided Communication Scheme for Downlink Coordinated Multipoint Transmission," *IEEE Access*, vol. 6, pp. 1416–1427, 2017.
- [45] W. Tan, M. Matthaiou, S. Jin, and X. Li, "Spectral Efficiency of DFT-Based Processing Hybrid Architectures in Massive MIMO," *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 586–589, 2017.
- [46] D. D. Nguyen, Y. Liu, and Q. Chen, "On the Energy Efficient Multi-Pair Two-Way Massive MIMO AF Relaying With Imperfect CSI and Optimal Power Allocation," *IEEE Access*, vol. 6, pp. 2589–2603, 2018.
- [47] Y. Liang, H. Wu, G. Huang, J. Yang, and H. Wang, "Thermal performance and service life of vacuum insulation panels with aerogel composite cores," *Energy and Buildings*, vol. 154, pp. 606–617, 2017.
- [48] J. Li, X. Jiang, Y. Yan, W. Yu, S. Song, and M. H. Lee, "Low Complexity Detection for Quadrature Spatial Modulation Systems," *Wireless Personal Communications*, vol. 95, no. 4, pp. 4171–4183, 2017.
- [49] J. Yang, H. Wu, S. He, and M. Wang, "Prediction of thermal conductivity of fiber/aerogel composites for optimal thermal insulation," *Journal of Porous Media*, vol. 18, no. 10, pp. 971–984, 2015.
- [50] J. Yang, H. Wu, M. Wang, S. He, and H. Huang, "Prediction and optimization of radiative thermal properties of ultrafine fibrous insulations," *Applied Thermal Engineering*, vol. 104, pp. 394–402, 2016.

