The SMART Handoff Policy for Millimeter Wave Heterogeneous Cellular Networks

Yao Sun, Gang Feng, *Senior Member, IEEE,* Shuang Qin, *Member, IEEE,*
Ying-Chang Liang, *Fellow, IEEE,* and Tak-Shing Peter Yum, *Fellow, IEEE,*

Abstract—The millimeter wave (mmWave) radio band is promising for the next-generation heterogeneous cellular networks (HetNets) due to its large bandwidth available for meeting the increasing demand of mobile traffic. However, the unique propagation characteristics at mmWave band cause huge redundant handoffs in mmWave HetNets that brings heavy signaling overhead, low energy efficiency and increased user equipment (UE) outage probability if conventional Reference Signal Received Power (RSRP) based handoff mechanism is used. In this paper, we propose a reinforcement learning based handoff policy named SMART to reduce the number of handoffs while maintaining user Quality of Service (QoS) requirements in mmWave HetNets. In SMART, we determine handoff trigger conditions by taking into account both mmWave channel characteristics and QoS requirements of UEs. Furthermore, we propose reinforcement-learning based BS selection algorithms for different UE densities. Numerical results show that in typical scenarios, SMART can significantly reduce the number of handoffs when compared with traditional handoff policies without learning.

Index Terms—Handoff, HetNets, Millimeter Wave, Reinforcement Learning.

1 Introduction

The 5th generation (5G) networks are expected to support the exponentially increasing demand of mobile traffic. A simple way to increase the network capacity is to allocate more bandwidth to 5G networks. Since the radio spectrum from 300MHz to 3GHz is very crowded, an effective solution is to design the 5G networks as two-tier heterogeneous cellular networks (HetNets) where the macrocell is supported by traditional cellular band, while some small or femto cells are supported by the globally available spectrum at millimeter wave (mmWave) band ranging from 30GHz to 300GHz [1]. This network architecture is called mmWave HetNets.

The key propagation properties at mmWave band are large propagation path loss and high sensitivity to blockage. These properties cause many design challenges for mmWave HetNets, including integrated circuits design, beamforming design, user association and handoff mechanisms. In particular, handoff is crucial for keeping users connected while moving around [1], [2]. Handoff mechanisms affect not only service quality of users but also network performance, such as throughput and energy efficiency. Conventional handoff mechanisms are based on Reference Signal Received Power (RSRP) measured by user equipments (UEs) [3]. In HetNets which are composed of traditional macro base stations and small base stations with low transmit power, handoff policies are mostly based on RSRP with Cell Range Expansion (CRE) [4]. With the introduction of mmWave band into cellular networks, it needs to co-exist with traditional communication bands forming a complex heterogeneous network [5]. Moreover, mmWave channel quality often changes rapidly and intermittently [6]. Therefore, using conventional handoff mechanisms in mmWave HetNets may lead to ping-pong effect, causing high outage probability and redundant handoffs.

Handoffs in mmWave HetNets are more frequent as mmWave cells are smaller. It was shown in [7] that the average handoff interval can be as low as 0.75 second in typical scenarios. A separate study [1] showed by computer simulation that more than 61% handoffs are unnecessary. The very large number of redundant handoffs causes heavy signaling overhead, low energy efficiency and high UE outage probability.

To reduce redundant handoffs in traditional HetNets, two parameters *hysteresis* and *threshold* are introduced in 3GPP [3]. For a specific UE, handoff is triggered if the RSRP of the current serving BS is lower than the *threshold* value and the RSRP of the target BS is stronger than that of the serving BS by *hysteresis*. This method, however, is not suitable for use in mmWave HetNets due to highly dense BSs deployments, small BS coverage and fast varying mmWave channel quality. There are also occasions where the "two-parameter" method misses *necessary* handoffs. Under these circumstances, artificial intelligence tools that incorporate information on surrounding environment can be used to design a smart handoff mechanism in mmWave HetNets.

In this paper, we propose a reinforcement learning based handoff policy named SMART for mmWave HetNets. Our design objective is to reduce the number of unnecessary handoffs while guaranteeing the QoS of UEs. SMART consists of two parts. Part 1 is to determine the handoff trigger condition by the mmWave channel characteristics and QoS requirements of UEs. Part 2 is on BS selections, and is carried out by two algorithms: SMART-S and SMART-M

This work was supported by the National Science Foundation of China under Grant number 61631005 and 61471089, and the Fundamental Research Funds for the Central Universities under Grant number ZYGX2015Z005.

- *Y. Sun, G. Feng, S. Qin and Y-C. Liang are with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, P.R. China. E-mail: fenggang@uestc.edu.cn*
- *T-S. Peter Yum is with College of Computer Science and Electronic Engineering, Hunan University, Changsha, P.R. China.*

for different UE density circumstances. SMART-S chooses target BS for single UE based on Upper Confidence Bound (UCB) algorithm that can achieve logarithmic performance when compared with the optimal algorithm that uses global perfect information. SMART-M is used for dense UE distribution circumstance to choose BSs for multiple UEs triggering handoffs in the same measurement report period. We formulate it as a 0-1 integer programming, and solve it by Lagrange dual decomposition with relaxation.

In the following, we introduce related works and the system model in Sections 2 and 3, respectively. In Section 4 we present the framework of our proposed handoff policy SMART. In Sections 5 and 6, we present the BS selection algorithm for a single and multiple UEs respectively. We compare the performance of SMART with traditional handoff policies in Section 7 and conclude the paper in Section 8.

2 Related Work

Here, we present the handoff policies for traditional HetNets and mmWave HetNets separately.

2.1 Handoff Strategies for Traditional HetNets

In recent years, research on handoff is focused mainly on HetNets operating in the band of traditional frequency 900MHz-2.4GHz and considering one or more factors including RSRP, QoS of UEs, UE mobility characteristics, BS load, etc. [8], [9], [10], [11], [12]. In [8], a handoff policy is proposed that considers context parameters, such as user speed, channel gains and cell load information. The BS selection decision is based on a Markov Decision Process (MDP) model with the aim of maximizing UE average capacity. In [9], the authors proposed a handoff algorithm based on BSs estimated load. They combined handoff decisions with BS sleeping policy so as to improve system energy efficiency. [10] and [11] are mainly focused on the improvement of handoff trigger conditions. The authors in [10] proposed a new handoff triggering mechanism named Network Controlled Handover (NCH) for 3GPP Long Term Evolution (LTE) HetNets. NCH can optimize handoff trigger parameters such as Channel Quality Indication (CQI) threshold based on the statistics of the handoff performance. In [11], the authors proposed a new handoff algorithm aiming at the efficient management of BSs transmitted power and the reduction of unnecessary handoffs. The authors of [12] proposed a novel handoff policy based on cooperation-based cell clustering in densely deployed HetNets to reduce handoff signaling overhead.

2.2 Handoff Strategies for mmWave HetNets

Thus far, there is little research work on handoff in mmWave HetNets [13], [14], [15], [16]. The authors of [13] proposed the Radio-over-Fiber (RoF) network architecture for mmWave communications which facilitates flexible and cost effective deployment of distributed antennas. They then proposed the Extended Cell (EC) in RoF architecture. An EC is a group of adjacent cells or antennas that transmit the same data over the same frequency channel for a specific UE. This can increase overlapping areas and thus decrease the outage probability of UE during handoffs. The method is suitable for indoor environment with low UE mobility. The authors of [14] proposed a dual connectivity (DC) network architecture to deal with the handoff between two radio access technologies (RATs): mmWave and LTE.

Focusing on the optimization of handoff policies for mmWave HetNets, the authors of [15] solved the BS selection problem by MDP through combining the contributions of handoff overhead, cell load and channel conditions into a reward function. The handoff policy can achieve high throughput while decreasing the number of handoffs. As the computation complexity of solving MDP is formidable, this strategy cannot readily applied to densely deployed HetNets. The authors of [16] developed an online learning-based approach to solve single UE network selection problem in heterogeneous wireless networks consisting of mmWave and other RATs, such as Wi-Fi and LTE. This work is focused on RAT selection for a single UE and aims at maximizing the long-term throughput of the UE. We will develop an approach in the following that reduces the number of unnecessary handoffs while guaranteeing the QoS of UEs. Besides, due to random line-of-sight mm-wave link, the authors of [17] suggest to assign more than one mmWave links to each user equipment so thus to decrease the signaling overhead for handoff in mmWave networks. They propose a joint access point placement and mobile device assignment scheme for mmWave networks with aim to minimize the number of access points while satisfying the line of sight coverage of mobile devices.

3 System Model

3.1 Network Scenario

Consider a densely deployed HetNet with M femto cells underlying a macrocell as shown in Fig. 1. Let $\mathcal{M}$ be the set of femto base stations (FBSs). FBSs can use either mmWave or the traditional cellular frequency shared with the macro base station (MBS). Let λ be the ratio of FBSs using mmWave frequency, $\mathcal{M}_m$ be the set of the mmWave FBS (denoted as mm-FBS), and $\mathcal{M}_t$ be the set of the traditional FBS (denoted as Tr-FBS). UEs move randomly in the HetNet.

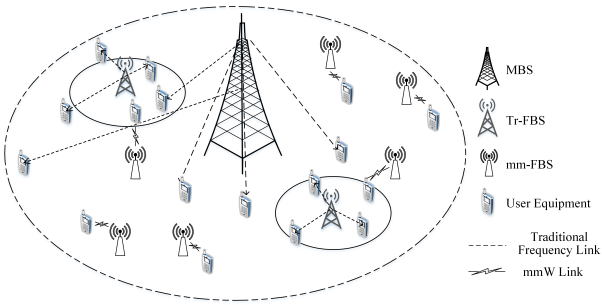


Fig. 1. The system model of mmWave HetNets.

3.2 Propagation Model

First, we discuss about the channel model of mmWave band. We assume that the channel of mm-FBS is based on 3GPP Standard probabilistic LOS-NLOS models [18], meaning that the channel condition between UE and mm-FBS can alternate between the two states, Line-of-Sight (LOS) and

Non-Line-of-Sight (NLOS). LOS state means that a line-of-sight mmWave link between UE and mm-FBS exists. The channel state transition probability is related to environment, and this probability is typically unknown [16]. Note that the channel state for UEs may be different, even when they are located at the same position and associated with the same mm-FBS. This is due to blockages, and thus the UEs may have different SNR. Similar to that in [5], [6], we assume that the path loss model is

$$L(d) = \alpha + 10\eta \log_{10}(d) + \xi\,[dB], \xi \sim N(0, \theta^2), \quad (1)$$

where d is the distance in meters, α and η are the least square fits of floating intercept and slope over the measured distances (30 to 200 m), and θ^2 is the lognormal shadowing variance. The values of α, η and θ are different for LOS and NLOS states [5], [6]. Since interference can be ignored for mm-FBS, for a specific UE, say UE n, the SNR when associated with mm-FBS j can be written as

$$SNR_n^j = \frac{P_j \psi L(d)^{-1}}{\sigma^2}, \quad (2)$$

where P_j is the transmit power of mm-FBS j, σ^2 is the noise power and ψ is the antenna gain. We assume that all mm-FBSs are equipped with directional antennas which are necessary to support beamforming and beam tracking for mmWave system. On the other hand, we assume that UEs are equipped with omnidirectional antennas, and thus the antenna gains are only accounted for at mm-FBSs side [19]. Similar to that in [5], [19], we assume that antenna gain model can be expressed as

$$\psi(\theta) = \begin{cases} \psi_{max},\ if\ |\theta| \le \frac{\theta_s}{2} \\ \psi_{min},\ otherwise \end{cases}, \quad (3)$$

where θ is the angle between UE and mm-FBS, and θ_s is the width of the antenna main lobe. When a UE is associated with an mm-FBS, in order to maintain the mmWave communication link, beam tracking could be used. We assume perfect beam tracking is performed, and thus the transmission direction of the UE is always in the main lobe, so as to enjoy a high antenna gain.

Next, we present the traditional radio band channel model. We assume that the MBS and Tr-FBSs are equipped with omnidirectional antennas to guarantee coverage area [19]. For traditional band links, we need to consider co-channel interference due to shared bandwidth deployment. The SINR of UE n associated with BS j can be expressed as

$$SINR_n^j = \begin{cases} \frac{P_j g_{nj}}{\sum_{k \in \mathcal{M}_t} P_k g_{nk} + \sigma^2},\ j \text{ is MBS} \\ \frac{P_j g_{nj}}{\sum_{k \in \{\mathcal{M}_t \cup MBS\}/\{j\}} P_k g_{nk} + \sigma^2},\ j \in \mathcal{M}_t \end{cases}, \quad (4)$$

where g_{nj} is the channel gain between UE n and BS j, which includes path loss and shadowing. Since we assume that Tr-FBSs and UEs are equipped with omnidirectional antennas, there is no extra antennas gain. For path loss, we use *flexible path loss exponent* model [20]

$$PL(d) = 10\epsilon \log_{10}(d) + 20 \log_{10} f + 32.45, \quad (5)$$

where d is the distance in meters, ϵ is the path loss exponent and f is the carrier frequency in MHz. For NLOS environments, a larger exponent is used [20]. For shadow fading, we use a zero mean Gaussian random variable ξ to describe it [21].

We assume that all BSs allocate bandwidth resources to their serving UEs uniformly. According to Shannon capacity formula, the achievable transmission rate for UE n associated with BS j can be written as

$$r_n^j = \begin{cases} \frac{B_m}{U_j} log_2(1 + SNR_n^j),\ j \in \mathcal{M}_m \\ \frac{B_t}{U_j} log_2(1 + SINR_n^j),\ j \in \{\mathcal{M}_t \cup MBS\} \end{cases}, \quad (6)$$

where B_m (B_t) is the bandwidth of mm-FBS (Tr-FBS and MBS) and U_j is the total number of UEs served by BS j.

3.3 Initial Access Model

In this subsection, we illustrate how to discover a new BS, and establish a possible connection in case a handover is performed. We assume that the cell search procedure *for traditional band* is identical to that in LTE, i.e. the MBS and Tr-FBSs perform cell search by transmitting omnidirectional synchronization signals [22]. *For mmWave system*, 4G-LTE initial access procedure is infeasible due to the problem of discovery range mismatch [23], [24]. In our model, we adopt an efficient initial cell search scheme, *iterative search* [25], [26], which performs a two-stage scanning procedure of the angular space. In detail, the space is partitioned into several wide sectors, and each wide sector is divided into several narrow sectors. In the first phase, the BS transmits pilots over wide sectors. In the second phase, the BS refines its search within the best wide sector by steering narrow beams, and thus finds the best narrow sector [23]. All the pilots are transmitted on a directional mmWave channel.

Technically, the cell search procedure is independent with the target BS selection policy. Hence, although the initial cell search scheme could affect the absolute value of the number of handoffs [27], [28], it does not affect the relative performance enhancement of the proposed SMART policy. Intuitively, some new cell search schemes, such as those proposed in [27] and [28] which use context information to speed up the cell search process, could be implemented with the proposed SMART handoffs policy in mmWave HetNets. As this is beyond the scope of the work, we use the aforementioned iterative search scheme for mmWave band initial cell search.

3.4 QoS Model

Similar to that in [29], [30], we use two factors to describe QoS requirement: minimum threshold of transmission rate γ_n^{min} and endurable time τ_n. The endurable time is the maximum time a UE is allowed to have the transmission rate lower than the minimum threshold. We state that the QoS of UE n is satisfied when the following condition holds

$$\exists t_0 \in [t - \tau_n, t], s.t. r_n^j(t_0) \ge \gamma_n^{min}. \quad (7)$$

Furthermore, to classify the type of service more precisely, we introduce a third factor: maximum threshold of transmission rate, denoted by γ_n^{max}. Let $\mathcal{C} = \{C_1, C_2, \cdots, C_L\}$ be the set of all service types, and specify that the service of UE n belongs to type C_i when $\tau_n \in [\tau_i, \tau_{i+1})$, $\gamma_n^{min} \in [\gamma_i^{min}, \gamma_{i+1}^{min})$ and $\gamma_n^{max} \in [\gamma_i^{max}, \gamma_{i+1}^{max})$. We assume that UEs in the system move at a random speed and in a random direction.

4 Framework of SMART Handoff Policy

3GPP Standard defines six handoff events for cellular networks [3] with Event A2 and Event A3 being the most common ones in HetNets. Our proposed SMART handoff policy focuses on these two handoff events, and other handoff decisions remain the same as those in 3GPP.

4.1 Handoff Trigger Conditions

Event A2 occurs when the RSRP of the serving BS becomes worse than a threshold [3], and the trigger condition can be expressed as

$$RSRP_n^j < threshold - Hys, \tag{8}$$

where *Hys* is a hysteresis parameter added for reducing redundant handoffs (e.g. ping-pong effect). Event A2 handoff is performed when the serving BS cannot fulfill the minimum UE QoS requirement. Thus, in SMART, the trigger condition can be written as

$$\forall t_0 \in [t-\tau_n, t], r_n^i(t_0) < \gamma_n^{min}, \tag{9}$$

where τ_n and γ_n^{min} are UE service type parameters. This change can avoid many unnecessary handoffs. Once inequality (9) is satisfied for UE n, an Event A2 handoff is triggered, and the UE needs to select a suitable target BS.

Event A3 occurs when a neighbor BS becomes *offset* better than the serving BS [3], and the trigger condition can be expressed as

$$RSRP_n^k \geq RSRP_n^j + offset, \tag{10}$$

for time-to-trigger (*TTT*) period, where $RSRP_n^k$ and $RSRP_n^j$ are the RSRPs of target BS k and current serving BS j measured by UE n respectively, and *offset* and *TTT* are two parameters defined in 3GPP. Once a UE experiences a handoff in this event, it means that the UE switches to a better BS which can improve its QoS although current serving BS can fulfill the minimum QoS requirement. Thus, SMART uses the following three trigger conditions

$$\exists t_0 \in [t-\tau_n, t], s.t. r_n^j(t_0) \geq \gamma_n^{min}, \tag{11-1}$$

$$r_n^k(t) \geq r_n^j(t) + offset, \tag{11-2}$$

$$\gamma_n^{max} - \gamma_n^{min} > \epsilon. \tag{11-3}$$

Condition (11-1) states that the current serving BS can fulfill the minimum UE QoS requirement. Condition (11-2) constraints that the transmission rate of the target BS k is at least *offset* higher than that of the serving BS j. Condition (11-3) indicates that the difference of transmission rate between maximum threshold and minimum threshold is greater than ϵ in QoS requirement. Similar to traditional handoff mechanism, when the above three conditions hold for *TTT* time, Event A3 handoff is triggered.

4.2 BS Selection

Once handoff trigger conditions are met, UEs need to select suitable target BSs. In SMART, we use reinforcement-learning for selecting BSs to reduce the number of unnecessary handoffs. We design two BS selection algorithms: SMART-S and SMART-M, for different UE density circumstances. SMART-S with low computational complexity is for a specific UE. It is suitable for sparse UE density circumstance. SMART-M is a joint optimal policy for multiple UEs triggering handoffs in the same measurement report period. It is suitable for dense UE distribution circumstance with a central controller.

5 SMART-S Algorithm for Single Target BS Selection

Note that once a specific BS satisfies the trigger conditions of Event A3, the target BS is determined. We therefore focus on the BS selection for Event A2. Let $\mathcal{A}_n(t)$ be the set of admissible BSs when UE n triggers Event A2 handoff at time t,

$$\mathcal{A}_n(t) = \{k \mid r_n^k(t) \geq \gamma_n^{min} + \Gamma, \forall k \in \mathcal{M} \cup MBS\}, \tag{12}$$

where Γ is a criteria offset parameter. For UE n with volume of data Q_n to be transmitted, we use H_n to denote the number of handoffs. Our goal is to select BS in set $\mathcal{A}_n(t)$ with minimum H_n once Event A2 condition is triggered.

5.1 Reinforcement-Learning Framework

We model the BS selection problem as a reinforcement learning problem. It consists of three elements: agent, environment and action. In our model shown in Fig.2, the agent is a specific UE n, the environment is the channel conditions of BSs, and the action is BS selection policy. The aim is to maximize the total reward by a sequence of BS selections.

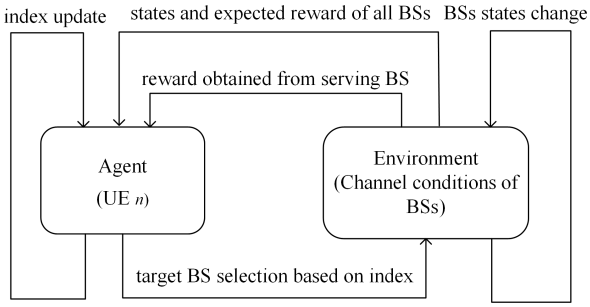


Fig. 2. Reinforcement learning based BS selection framework

Our objective is to minimize the total number of handoffs H_n. As it is difficult to incorporate H_n into the reward function directly, we make a transformation as follows. Let reward function $R_n^k(t)$ be defined as the volume of transmitted data from time t to t_n^k when UE n switches to BS k at time t, or

$$R_n^k(t) = \int_t^{t_n^k} r_n^k(t)dt. \tag{13}$$

Proposition 1. *Minimizing the total number of handoffs H_n for UE n is equivalent to solving the proposed reinforcement learning problem with the reward function defined in (13).*

Proof: Let t_n^k in (13) equal to the time when the next handoff for UE n is triggered after time t, and we define a sort function Φ in a finite set X as

$$\Phi(x) = k, x \in X \text{ and } x \text{ is the } k \text{ smallest element in } X. \tag{14}$$

The objective of the above reinforcement learning model is to find the optimal policy π^*:

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\pi}[\sum_{\Phi(t_n^k)=1}^{K} R_n^k(t)], \tag{15}$$

where K is the maximum value of $\Phi(t_n^k)$, which is equals to the number of handoffs in the time period.
If we fix the volume of transmitted data of UE n as Q_n, applying policy π^* can minimize the total number of handoffs of UE n when transmitting Q_n data, which equals to our optimization objective $\min H_n$. □

5.2 Expected Reward Estimation

As t_n^k and $r_n^k(t)$ in (13) are unknown random variables, the expected reward $\mathbb{E}[R_n^k(t)]$ can only be estimated from historical information. We use $\check{R}_n^k(t)$ to denote the observed value of $R_n^k(t)$ which can be obtained once UE n switches to BS k. However, a UE may not stay around a specific BS k for a long time, and thus we cannot have enough historical information to estimate $R_n^k(t)$ accurately. To get around, we define *type reward* $\check{R}_{C_n}^k(T_{C_n}^k)$ as

$$\check{R}_{C_n}^k(0) = 0, \tag{16-1}$$

$$\check{R}_{C_n}^k(T_{C_n}^k + 1) = \frac{T_{C_n}^k \check{R}_{C_n}^k(T_{C_n}^k) + \check{R}_n^k(t)}{T_{C_n}^k + 1}, \tag{16-2}$$

where $T_{C_n}^k$ denotes the number of times that BS k is selected by UEs with service type C_n. We take this observed value $\check{R}_{C_n}^k(T_{C_n}^k)$ as the mean reward for UEs with the same service type C_n, and each UE uses his own observed reward $\check{R}_n^k(t)$ to update the type reward $\check{R}_{C_n}^k(T_{C_n}^k)$ after a handoff occurs based on (16-2). Thus, the expected reward can be estimated as

$$\mathbb{E}[R_n^k(t)] = \begin{cases} \check{R}_{C_n}^k(T_{C_n}^k), & \text{if } n \in C_n \\ 0, & \text{otherwise} \end{cases}. \tag{17}$$

Since the handoff trigger conditions of UEs with the same service type are similar, type reward $\check{R}_{C_n}^k(T_{C_n}^k)$ can be accurately estimated by reinforcement learning.

5.3 BS Selection Algorithm

We cannot always select the BS with the highest reward since a well-known dilemma exploration vs. exploitation exists in reinforcement learning. This dilemma states that there is a tradeoff between improving UEs knowledge about the reward distributions of BSs (exploration) and switching to the BS with the highest empirical mean reward (exploitation). Regret is a concept to measure the performance of a policy [16], which is defined as the difference of total reward between the adopted policy and global optimal policy. In our problem, the regret of policy π after W handoffs occur for UE n can be expressed as

$$Regret_{\pi}(W) = \sum_{\Phi(t_n^k)=1}^{W} [R_{\pi^*}(t_n^k) - R_n^k(t_n^k)_{\pi}], \tag{18}$$

where $R_{\pi^*}(t_n^k)$ is the reward of the optimal policy π^* at time t_n^k. It was shown in [31] that the best regret is logarithmic with respect to the number of handoffs W. Based on that, the authors of [32] proposed an Upper Confidence Bound (UCB) algorithm to deal with this tradeoff. It can achieve logarithmic regret with low computation complexity. The UCB policy states that the agent chooses machine j^* at each decision time according to the following index

$$j^* = \arg\max_{j} \left(\overline{x}_j + \sqrt{\frac{2 \ln W}{W_j}} \right), \tag{19}$$

where $\overline{x}_j$ is the average reward obtained from machine j, W_j is the number of times machine j has been chosen and W is the overall number of decisions so far.

The BS selection algorithm when UE n triggers Event A2 handoffs is based on UCB. We set index of BS j for UE n as $\check{R}_{C_n}^k(T_{C_n}^k) + \ell \sqrt{\frac{2 \ln H_n}{T_{C_n}^k}}$, where $\ell = \max_{k \in \mathcal{A}_n, C_n \in \mathcal{C}} \check{R}_{C_n}^k(T_{C_n}^k)$ and H_n is the total number of handoffs for UE n so far. Thus, the policy is selecting BS k^* in set $\mathcal{A}_n$ for UE n once Event A2 handoff occurs, where k^* can be expressed as

$$k^* = \arg\max_{k} \left(\check{R}_{C_n}^k(T_{C_n}^k) + \ell \sqrt{\frac{2 \ln H_n}{T_{C_n}^k}} \right). \tag{20}$$

We summarize our proposed SMART-S BS selection algorithm in Algorithm 1.

Algorithm 1 : SMART-S BS selection algorithm based on UCB.

Input: Network topology (BS and UE distributions, λ); service type of UEs.
Output: BS selection decisions k^*.

```
Initialization: obtain T_{C_n}^k, H_n, Ř_{C_n}^k(T_{C_n}^k) in time T based on traditional handoff policy
while handoff conditions are met for a certain UE n do
    if Event A2 handoff then
        Judge service type C_n of UE n
        Ř_{C_n}^k(T_{C_n}^k + 1) ← (T_{C_n}^k Ř_{C_n}^k(T_{C_n}^k) + Ř_n^k(t)) / (T_{C_n}^k + 1)
        k* = arg max_k ( Ř_{C_n}^k(T_{C_n}^k) + ℓ sqrt(2 ln H_n / T_{C_n}^k) )
        T_{C_n}^k ← T_{C_n}^k + 1, H_n ← H_n + 1
    else
        switch to the unique target BS k*
    end if
end while
```

5.4 Properties of SMART-S

SMART-S algorithm does not perform iteration and thus does not have convergence issue. We investigate here the performance bound and signaling overhead. The performance bound is established from Fact 1 and Corollary 1.

Fact 1. For all $K > 1$, if policy UCB is run on K machines having arbitrary reward distributions $P_1, \cdots, P_K$ with support in $[0, 1]$, its expected regret can achieve logarithmic bound.

Proof: cf. [32] for proofs. □

Corollary 1. The proposed UCB-based SMART-S BS selection policy achieves logarithmic regret with respect to the total number of handoffs H_n.

Proof: We construct a new reinforcement learning model which is the same as our above proposed model except for the reward function. For sake of convenience, we denote the above proposed and the new reinforcement learning model as RL 1 and RL 2 respectively. The reward function of RL 2 is defined as

$$Y_n^k(t) = \frac{R_n^t(t)}{\ell}, \tag{21}$$

where $R_n^k(t)$ is the reward function of RL 1 defined in (13). As ℓ is a constant, RL1 and RL2 have the same policy solution. Since $Y_n^k(t)$ has a bounded support in $[0,1]$, we use UCB algorithm to solve RL 2 problem, and thus the index in (19) can be expressed as

$$k^* = \arg\max_k \left(\overline{y}_k + \sqrt{\frac{2\ln H_n}{T_{C_n}^k}} \right), \tag{22}$$

where $\overline{y}_k$ is the average reward obtained from BS k which equals to $\frac{\check{R}_{C_n}^k(T_{C_n}^k)}{\ell}$. Thus, the index can be rewritten as

$$k^* = \arg\max_k \left(\frac{\check{R}_{C_n}^k(T_{C_n}^k)}{\ell} + \sqrt{\frac{2\ln H_n}{T_{C_n}^k}} \right). \tag{23}$$

According to Fact 1, we know that the regret bound is logarithmic for RL 2 problem, and thus for RL 1. Since ℓ is a constant, we use

$$k^* = \arg\max_k \left(\check{R}_{C_n}^k(T_{C_n}^k) + \ell\sqrt{\frac{2\ln H_n}{T_{C_n}^k}} \right) \tag{24}$$

to replace the index in (23) which is the same as the proposed BS selection policy. □

Next we discuss the signaling overhead for SMART-S BS selection algorithm. When the handoff trigger conditions are satisfied for a specific UE, it notifies his service type to the admissible BSs in set $\mathcal{A}_n$, and the BSs calculate and send their corresponding indexes to the UE. The UE switches to the target BS, say BS k, determined by using (20). When the next handoff occurs, the UE obtains the value of $\check{R}_n^k(t)$, and transmits it to BS k. The BS uses it to update the expected reward and index according to (16) and (17). Thus, the number of signaling exchanges needed is $2|\mathcal{A}_n|$ and each signaling exchange uses several bits.

6 SMART-M Algorithm for Multiple Target BS Selection

The BS selection algorithm discussed in Section 5 focuses on individual UEs. However, in the time interval between two adjacent measurement report periods, there may be multiple UEs that need handoff especially for dense UE distribution. Moreover, multiple UEs may trigger handoffs in the same time period or even simultaneously in typical scenarios, such as a group of UEs riding in a moving bus. We therefore design SMART-M algorithm for optimal multi-BS selection.

6.1 Problem Formulation based on Learning Results

Let $\mathcal{N}$ be the set of UEs sending handoff request to the network central controller in a measurement period and let $N = |\mathcal{N}|$. As the period is usually short (e.g. in tens of milliseconds), we assume that the BS selection decisions are made at the end of individual periods. Here, the objective function Y is again chosen as the volume of transmitted data before the next handoff occurs for these N UEs. Also we use $\check{R}_{C_i}^j(T_{C_i}^j)$ to estimate $E[R_n^k(t)]$ based on the above reinforcement learning. The problem is formulated as

$$\max\ Y = \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{A}_i} x_{ij}\check{R}_{C_i}^j(T_{C_i}^j) \tag{25}$$

$$\text{s.t.} \sum_{i\in\mathcal{N}} x_{ij} \le N_j, \forall j \in \cup_{i\in\mathcal{N}}\mathcal{A}_i, \tag{25-1}$$

$$\sum_{j\in\mathcal{A}_i} x_{ij} = 1, \forall i \in \mathcal{N} \tag{25-2}$$

$$x_{ij} \in \{0,1\}, \forall i \in \mathcal{N}, \forall j \in \cup_{i\in\mathcal{N}}\mathcal{A}_i, \tag{25-3}$$

where x_{ij} is a binary variable indicating whether UE i switches to BS j, N_j is the current connection capacity of BS j (equals to the maximum connection capacity minus the number of current serving UEs), and $\mathcal{A}_i$ is the set of admissible BSs for UE i. Constraint (25-1) ensures that the number of UEs which switch to the same BS does not exceed the current BS connection capacity. Constraints (25-2) and (25-3) guarantee that each UE can only be associated with one BS at a time. For convenience, we use set $\mathcal{A}$ to denote $\cup_{i\in\mathcal{N}}\mathcal{A}_i$ in the rest of the paper.

6.2 BS Selection

The problem stated in (25) is a special case of a well-known NP-hard problem Generalized Assignment Problem (GAP), with $O(N^{|\mathcal{A}|})$ complexity using brute force algorithm. Obviously it is infeasible to use the brute force algorithm for solving dense deployment mmWave HetNets due to prohibitively high computational complexity. Instead, we propose the following efficient heuristics. We first relax binary variables x_{ij} in constraints (25-3) to be continuous variables in $[0,1]$. We then exploit Lagrange dual decomposition method [33] to solve this optimization problem.

After relaxing x_{ij}, problem (25) becomes a linear problem with Lagrange function

$$L(\boldsymbol{x},\boldsymbol{\mu}) = \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{A}_i} x_{ij}\check{R}_{C_i}^j(T_{C_i}^j) - \sum_{j\in\mathcal{A}}\mu_j\Big(\sum_{i\in\mathcal{N}} x_{ij} - N_j\Big), \tag{26}$$

where μ_j is Lagrange multiplier. For a fixed vector $\boldsymbol{\mu}$, Lagrange dual function can be expressed as

$$g(\boldsymbol{\mu}) = \sup_{\boldsymbol{x}}\ L(\boldsymbol{x},\boldsymbol{\mu}) \tag{27}$$

$$\text{s.t.} \sum_{j\in\mathcal{A}_i} x_{ij} = 1, \forall i \in \mathcal{N}, \tag{27-1}$$

$$0 \le x_{ij} \le 1, \forall i \in \mathcal{N}, \forall j \in \mathcal{A}, \tag{27-2}$$

and the dual problem is $\min_{\boldsymbol{\mu}} g(\boldsymbol{\mu})$. Rewriting function $g(\boldsymbol{\mu})$ yields

$$g(\boldsymbol{\mu}) = \sup_{\boldsymbol{x}} \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{A}_i} x_{ij}(\check{R}_{C_i}^j(T_{C_i}^j) - \mu_j) + \sum_{j\in\mathcal{A}}\mu_j N_j. \tag{28}$$

Since it dose not include the cross-term of x_{ij}, we can exchange the computation order as:

$$g(\boldsymbol{\mu}) = \sum_{i\in\mathcal{N}} \sup_{x_{ij}, j\in\mathcal{A}_i} \sum_{j\in\mathcal{A}_i} x_{ij}(\check{R}^j_{C_i}(T^j_{C_i}) - \mu_j) + \sum_{j\in\mathcal{A}} \mu_j N_j. \tag{29}$$

Thus, we can solve the following problem for each UE i separately,

$$g_i(\boldsymbol{\mu}) = \sup_{x_{ij}, j\in\mathcal{A}_i} \sum_{j\in\mathcal{A}_i} x_{ij}(\check{R}^j_{C_i}(T^j_{C_i}) - \mu_j) \tag{30}$$

$$\text{s.t.} \sum_{j\in\mathcal{A}_i} x_{ij} = 1, \tag{30-1}$$

$$0 \le x_{ij} \le 1. \forall j \in \mathcal{A}_i. \tag{30-2}$$

Since we want to find a binary solution of x_{ij}, for a fixed vector $\boldsymbol{\mu}$, problem (30) is described as: for UE i, we choose a BS j^* from set $\mathcal{A}_i$ to maximize the value of $R^{j^*}_{C_i}(t) - \mu_{j^*}$. Therefore, when $\boldsymbol{\mu}$ is fixed, problem (27) can be solved by choosing the optimal BS j^* for each UE respectively. Then we minimize $g(\boldsymbol{\mu})$ over $\boldsymbol{\mu}$ to obtain the optimal value $\boldsymbol{\mu}^*$ for the dual problem. We use negative gradient direction to update μ_j with respect to $\mu_j \ge 0$,

$$\mu_j(k+1) = \left[\mu_j(k) - \delta(k)(N_j - \sum_{i\in\mathcal{N}} x_{ij})\right]^+, \forall j \in \mathcal{A}, \tag{31}$$

where $\delta(k) > 0$ is the update step size, and is given by

$$\delta(k) = \frac{g(\boldsymbol{\mu}_k) - g_k}{\|\boldsymbol{h}_k\|^2}, \forall k \ge 0, \tag{32}$$

where g_k is an estimate of the optimal value g^*. The procedure of updating g_k is given by

$$g_k = \min_{1\le j\le k} g(\boldsymbol{\mu}_k) - \varepsilon_k, \tag{33}$$

and ε_k is updated according to

$$\varepsilon_{k+1} = \begin{cases} \rho\varepsilon_k & \text{if } g(\boldsymbol{\mu}_{k+1}) \le g_k \\ \max\{\beta\varepsilon_k, \varepsilon\} & \text{otherwise} \end{cases}, \tag{34}$$

where ε, β and ρ are fixed positive constant with $\beta < 1$ and $\rho \ge 1$ [34].

For linear programs, strong duality holds. Therefore, the minimum value of $g(\boldsymbol{\mu})$ is equal to the maximum value of the original problem. The solution process is that: we first obtain the maximum value $g(\boldsymbol{\mu})$ over $\boldsymbol{x}$ with fixed $\boldsymbol{\mu}$, and then minimize $g(\boldsymbol{\mu})$ over $\boldsymbol{\mu}$ denoted as $g(\boldsymbol{\mu}^*)$. The optimal binary solution $\boldsymbol{x}^*$ is obtained with the corresponding solution $\boldsymbol{\mu}^*$. According to $\boldsymbol{x}^*$ we make BS selections for those handoff UEs.

Similar to that in Section 5, $\check{R}^j_{C_i}(T^j_{C_i})$ is updated once the next handoff occurs according to (14) and (15). Note that, the reinforcement-learning process in Section 5 can improve the accuracy of the value of $\check{R}^j_{C_i}(T^j_{C_i})$ thus the solution of this optimization problem. We summarize the SMART-M algorithm in Algorithm 2.

6.3 Convergence and Computational Complexity of SMART-M

To prove the convergence of SMART-M BS selection algorithm, we need Propositions 2 and 3.

Algorithm 2 : Joint optimal SMART-M BS selection algorithm.

Input: Network topology (BS and UE distributions, λ); handoff UEs $\mathcal{N}$.
Output: BS selection decisions $\boldsymbol{x}^*$.

Initialization:
1: Judge service type of UEs
2: Determine admissible BSs
3: The BSs send the value of $\check{R}^j_{C_i}(T^j_{C_i})$ and N_j to the central controller
BS selection decisions:
4: $\boldsymbol{x}^0 \leftarrow 0$, $\boldsymbol{x}^0 \leftarrow$ current connections, $k \leftarrow 1$
5: **while** $\boldsymbol{x}^k \ne \boldsymbol{x}^{k-1}$ **do**
6: $k \leftarrow k+1$
7: **for** each UE $i \in \mathcal{N}$ **do**
8: solve problem (30)
9: **end for** (obtain $\boldsymbol{x}^k$)
10: update $\boldsymbol{\mu}^k$ according to (31)
11: **end while**
12: $\boldsymbol{x}^* \leftarrow \boldsymbol{x}^k$

Proposition 2. *Let $\boldsymbol{\mu}_k$ be the sequence generated by (31). Then, for all non-negative $|\mathcal{A}|$-dimensional vectors $\boldsymbol{v}$ and $k \ge 0$*

$$\|\boldsymbol{\mu}_{k+1} - \boldsymbol{v}\|^2 \le \|\boldsymbol{\mu}_k - \boldsymbol{v}\|^2 - 2\delta(k)(g(\boldsymbol{\mu}_k) - g(\boldsymbol{v})) + \delta(k)^2\|\boldsymbol{h}(\boldsymbol{x}_k)\|^2,$$

where $\|\boldsymbol{h}(\boldsymbol{x}_k)\|$ is an $|\mathcal{A}|$-dimensional vector with elements $h_j = N_j - \sum_{i\in\mathcal{N}} x^k_{ij}$, $|\mathcal{A}|$ is the cardinality of set $\mathcal{A}$, and $\boldsymbol{x}_k$ is a vector that satisfies $L(\boldsymbol{x}_k, \boldsymbol{\mu}_k) = \sup_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\mu}_k) = g(\boldsymbol{\mu}_k)$.

Proof: According to (31), we have

$$\begin{aligned}\|\boldsymbol{\mu}_{k+1} - \boldsymbol{v}\|^2 &= \|\boldsymbol{\mu}_k - \delta(k)\boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{v}\|^2 \\ &= \|\boldsymbol{\mu}_k - \boldsymbol{v}\|^2 - 2\delta(k)(\boldsymbol{\mu}_k - \boldsymbol{v})^T\boldsymbol{h}(\boldsymbol{x}_k) + \delta(k)^2\|\boldsymbol{h}(\boldsymbol{x}_k)\|^2.\end{aligned} \tag{35}$$

As $L(\boldsymbol{x}, \boldsymbol{\mu})$ in (26) is linear and $\boldsymbol{h}(\boldsymbol{x}) = \frac{\partial L(\boldsymbol{x},\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}$, for all vectors $\boldsymbol{x} \in [0,1]^{|\mathcal{N}|\times|\mathcal{A}|}$ we have

$$L(\boldsymbol{x}, \boldsymbol{\mu}_k) - L(\boldsymbol{x}, \boldsymbol{v}) = (\boldsymbol{\mu}_k - \boldsymbol{v})^T\boldsymbol{h}(\boldsymbol{x}). \tag{36}$$

For vector $\boldsymbol{x}_k$,

$$\begin{aligned}L(\boldsymbol{x}_k, \boldsymbol{\mu}_k) - L(\boldsymbol{x}_k, \boldsymbol{v}) &= g(\boldsymbol{\mu}_k) - L(\boldsymbol{x}_k, \boldsymbol{v}) \\ &\ge g(\boldsymbol{\mu}_k) - \sup_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{v}) = g(\boldsymbol{\mu}_k) - g(\boldsymbol{v}).\end{aligned} \tag{37}$$

Combining (36) and (37) yields

$$(\boldsymbol{\mu}_k - \boldsymbol{v})^T\boldsymbol{h}(\boldsymbol{x}) \ge g(\boldsymbol{\mu}_k) - g(\boldsymbol{v}). \tag{38}$$

Combining (35) and (38) yields

$$\|\boldsymbol{\mu}_{k+1} - \boldsymbol{v}\|^2 \le \|\boldsymbol{\mu}_k - \boldsymbol{v}\|^2 - 2\delta(k)(g(\boldsymbol{\mu}_k) - g(\boldsymbol{v})) + \delta(k)^2\|\boldsymbol{h}(\boldsymbol{x}_k)\|^2.$$

□

Proposition 3. *Assuming that step size $\delta(k)$ is determined by (32), (33) and (34), if $g^* > -\infty$ then $\lim_{k\to\infty} \inf g(\boldsymbol{\mu}_k) \le g^* + \varepsilon$.*

Proof: As N_j and x_{ij}^k are bounded, $h_j = N_j - \sum_{i\in\mathcal{N}} x_{ij}^k$ and then vector $\boldsymbol{h}(\boldsymbol{x}_k)$ are also bounded. Since $\boldsymbol{h}(\boldsymbol{x}) = \frac{\partial L(\boldsymbol{x},\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}$, there exists a scalar c that

$$c \geq \sup\{\|\partial g(\boldsymbol{\mu})\|\}. \tag{39}$$

Combining Proposition 2 and (39), we can conclude that our problem satisfies the necessary conditions of Proposition 6.3.6 in [34]. By applying this proposition with $\gamma_k = 1$, we obtain the Proposition 3, and the convergence of SMART-M is proved. □

Next we discuss the computational complexity of SMART-M BS selection algorithm. At each iteration, we decompose $g(\boldsymbol{\mu})$ into N sub-problems $g_i(\boldsymbol{\mu})$, and the computational complexity of $g_i(\boldsymbol{\mu})$ is $O(|\mathcal{A}_i|)$. Thus, the complexity of SMART-M BS selection algorithm is $O(k|\mathcal{N}||\mathcal{A}|)$, where k is the number of iterations. In most simulation experiments, the algorithm converges in less than 10 iterations with total run time several milliseconds, which can satisfy real-time requirements.

We again evaluate the signaling overhead for SMART-M BS selection algorithm. The UEs who trigger handoff conditions need to notify the service types to their admissible BSs, which calculate and send the corresponding value $\check{R}_{C_i}^j(T_{C_i}^j)$ to the central controller. The central controller makes handoff decisions based on SMART-M policy, and sends these decisions to UEs. Thus, the number of signaling exchanges needed is $\sum_{i=1}^N |\mathcal{A}_i|+|\mathcal{A}|+|\mathcal{N}|$, and each signaling exchange uses several bits.

6.4 Further Discussions on SMART-S and SMART-M

After discussing the details of the two algorithms separately, in this subsection, we clarify the two algorithms together from two aspects: (1) the relation between the two algorithms; and (2) the implementation of the algorithms.

First, the two algorithms are designed for different UE density scenarios. SMART-S is appropriate for sparse UE density, while SMART-M is designed for dense UE distribution. Specifically, SMART-S chooses target BS for a single UE without considering the states and decisions of other UEs. SMART-M can achieve a joint optimal BS selection policy for multiple UEs which are triggered to perform handoffs in the same measurement report period. The computational complexity of SMART-M is higher than SMART-S algorithm. Thus, we choose SMART-S or SMART-M according to the UE density.

On the other hand, SMART-M makes handoff decisions for multiple UEs by solving an optimization problem with unknown parameters. We employ the learning algorithm of SMART-S to evaluate the unknown parameters in the optimization framework. In more details, SMART-M needs to solve (25) with the expected reward $\check{R}_{C_i}^j(T_{C_i}^j)$ in the reinforcement learning model of SMART-S, which is indeed the estimated value of $\mathbb{E}[R_n^k(t)]$. In this sense, SMART-S can be used to enhance the accuracy of $\check{R}_{C_i}^j(T_{C_i}^j)$ from historical data, and thus improve the performance of SMART-M.

Second, let us discuss the implementations of the two algorithms, in order to further clarify their relation. For the selection between the two algorithms, SMART-S is indeed feasible for any UE density. From Corollary 1, we can see that SMART-S achieves logarithmic regret with respect to the total number of handoffs. In other words, although we always run SMART-S for any UE density, we can still achieve at least logarithmic regret bound and enjoy performance improvement. Certainly, for dense UE distribution circumstances, we can run SMART-M algorithm to further improve the handoff performance with some computational cost.

In our system, we adopt a simple selection policy between the two algorithms. We define a UE density threshold Γ to identify sparse or dense UE distribution. When the number of UEs which send handoff request to the controller in a measurement period is lower than Γ, SMART-S is selected, otherwise SMART-M is selected. Other handoff procedure remains the same as that in conventional handoff policy.

7 Numerical Results

In this section, we compare the performance of SMART with two conventional handoff policies as follows. (1) Rate-based handoff (RBH). RBH has similar trigger conditions as those in 3GPP. When choosing target BSs for handoffs, the ones with maximum transmission data rates are chosen (instead of maximum RSRP in 3GPP [3]). (2)SINR based handoff (SBH). SBH has the same handoff trigger conditions as that of SMART and uses maximum SINR for target BS selection.

7.1 Simulation Settings

We consider a two-tier HetNet deployed in urban area, and the HetNet consists of an MBS and varying number of mm-FBSs, Tr-FBSs and UEs. The MBS is located at the central of a circular area with radius equal to 500m, and both mm-FBSs and Tr-FBSs are randomly distributed in the area. The transmit power of MBS, mm-FBS and Tr-FBS is set to 46dBm, 30dBm and 20dBm, respectively. Both the number and region of blockages in mm-FBS are randomly generated. Similar to that in [6], when UEs in mm-FBS move to blockage regions, the channel state is assumed to be NLOS with parameters$\alpha = 72$ and $\eta = 2.92$ in (1). In non-blockages areas, the channel state is assumed to be LOS with parameters $\alpha = 61.4$ and $\eta = 2$ in (1). Other parameters related to mmWave band path loss model are the same as those in [6]. For traditional band, the carrier frequency is set to 2GHz, and we use path loss model in (5) with different exponent, $\epsilon = 2$ for LOS and $\epsilon = 3$ for NLOS. The bandwidth allocated to MBS/Tr-FBSs and mm-FBSs is 20MHz and 500MHz respectively. The noise power is set to -101dBm and -77dBm for traditional and mmWave band respectively [5]. We assume that the UEs are randomly distributed in the area and move to a random direction at a random speed. We assume that perfect initial cell search can be performed, and thus UEs can discover BSs correctly when a handoff occurs. TABLE I summarizes the system parameters we use in the simulations. Our simulations are implemented with MATLAB codes and carried out on a PC equipped with an Intel-i5 4 core 3.2GHz processor and 4G RAM.

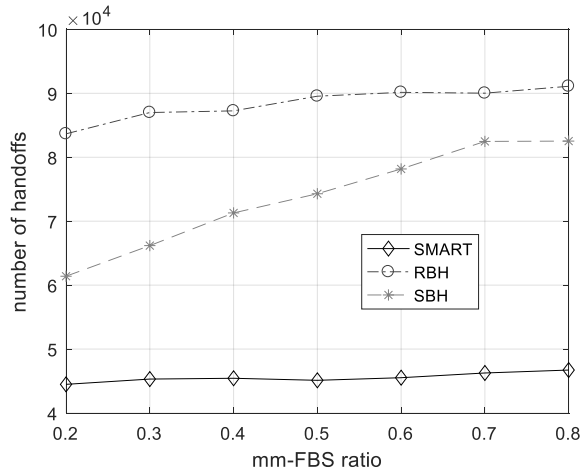
TABLE 1
Simulation Parameters

Parameters	Value
MBS radius	500 m
Power of MBS	46dBm
Power of mm-FBSs	30dBm
Power of Tr-FBSs	20dBm
Bandwidth of MBS/Tr-FBS	20 MHz
Bandwidth of mm-FBS	500 MHz
Path loss exponent for LOS	2
Path loss exponent for NLOS	3
ψ_{max}	18dB
ψ_{min}	-2dB
Carrier frequency	2000 MHz
Parameters for LOS path loss	$\alpha = 72$; $\eta = 2.92$
Parameters for NLOS path loss	$\alpha = 61.4$; $\eta = 2$
Noise power for mmWave band	-77 dBm
Noise power for traditional band	-101 dBm

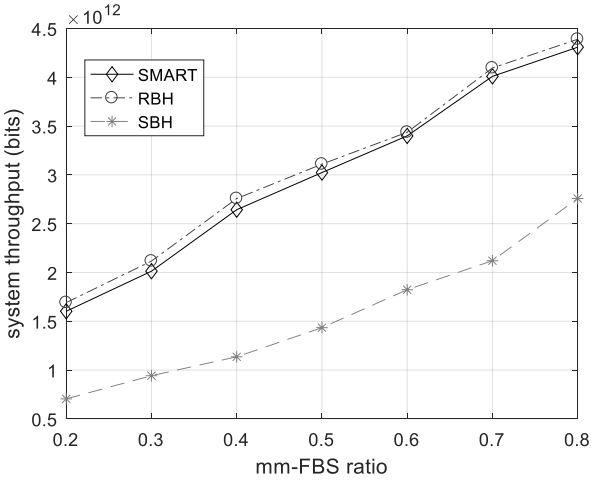
7.2 Numerical Results and Discussions

In Experiment 1, we compare the number of handoffs and system throughput of the three handoff policies. In this experiment, we fix the number of FBSs and UEs as 100 and 500 respectively. The average UE movement speed is 5m/s. Fig.3 shows the number of handoffs and system throughput for the three handoff policies with different mm-FBS ratio λ in 1000 seconds. Fig.3 (a) shows that when $\lambda = 0.2$, the total number of handoffs for RBH, SBH and SMART is 8.3×10^4, $6.1 \times 10^4$and 4.4×10^4, respectively. These numbers show that SMART can reduce handoffs by 47% and 28% when compared with RBH and SBH respectively. For $\lambda = 0.8$, the reduction percentages are 50% and 46%. Note that fewer handoffs implies reduced signaling overhead, energy consumption and UE outage probability. Fig.3 (b) shows that the system throughput of all the three handoff policies increases with the ratio of mm-FBS because of increasing available bandwidth in mm-FBS. The system throughput of RBH is higher than that of the other two schemes since that the handoff trigger conditions in RBH takes into account only UE data rate. In other words, in RBH a UE may frequently perform handoff for achieving maximum data rate, while ignoring the negative effective of handoff. We also find that the difference of system throughput between SMART and RBH is relatively small (2% for $\lambda = 0.8$, 5% for $\lambda = 0.2$), implying that significant handoff performance gain can be accomplished with a small compromise on throughput.

In Experiment 2, we evaluate the average running time per handoff (RT) of the three handoff policies with varying number of FBSs. The simulation settings used remain the same with those in Experiment 1. RT directly reflects the computational complexity for a handoff policy. Fig.4 shows the RT of the three handoff policies as a function of the number of FBSs. From the figure, we can see that the run-



(a) the number of handoffs



(b) system throughput

Fig. 3. Handoff performance as a function of mm-FBS ratio λ

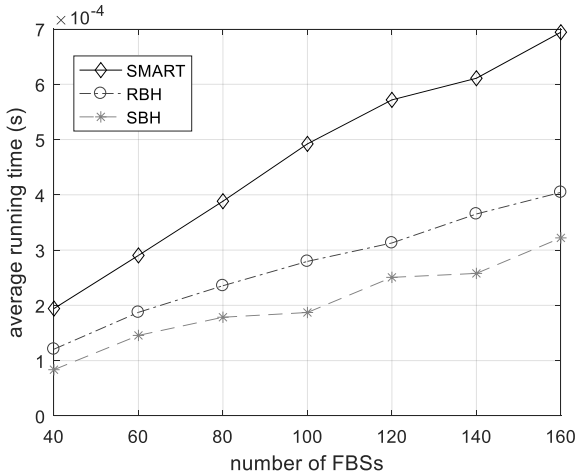


Fig. 4. Running time comparisons for handoff policies.

ning time of SMART increases approximately linearly with the number of FBSs. Moreover, we can see that although the RT of SMART is always the largest, it is still in the same order of magnitude (within 2-3 times) as that of the other two policies.

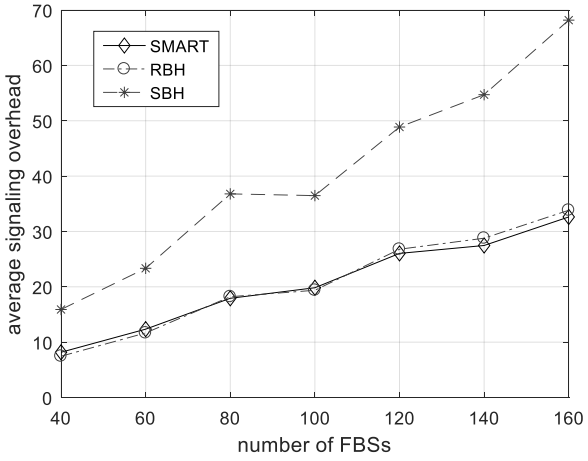


Fig. 5. Signaling overhead comparisons for handoff policies.

In Experiment 3, we examine the average signaling overhead (SOH), which is defined as the number of signaling exchanges per handoff with varying number of FBSs by using the same simulation settings. Fig.5 shows the SOH of the three handoff policies as a function of the number of FBSs. Note that in the experiment, SMART-S or SMART-M algorithm is selected according to the number of handoff UEs in each measurement report period, and we count the total number of signaling exchanges in 1000s time as the SOH of SMART. From the analysis in Section V. (D) and Section VI. (C), we know that the SOH of both SMART-S and SMART-M increases with the number of FBSs in a linear fashion with different slope. From the figure, we can observe that trend of SOH for SMART increases approximately linearly with the number of FBSs, which is concordant with the theoretical analysis. Moreover, we find that the curve of SMART is very close to that of RBH, which means that SMART handoff policy does not introduce additional signaling overhead. This is because that almost all the handoff procedures of SMART remain the same as that in conventional handoff policy, except for reward update and handoff algorithm selection.

In Experiment 4, we examine the effect of UE movement speed at $\lambda = 0.5$ with parameters the same as the Experiment 1. Fig.6 shows the number of handoffs and system throughput for the three handoff policies as a function of the mean UE movement speed. From Fig. 6 (a), we see that from fast walking speed of 2 m/s (7.2 km/h) to slow driving of speed of 14 m/s (50km/h), the numbers of handoffs are increased slightly for all three policies. The relative advantage of SMART remains. As expected, Fig.6 (b) shows that the system throughput of all the three policies decreases with UE movement speed due to faster change of channel quality.

In Experiment 5, we examine the performance of handoff policies for varying number of FBSs while using fixed mm-FBS ratio 0.5. Other parameters remain the same as those of the Experiment 1. Fig. 7 shows the number of handoffs

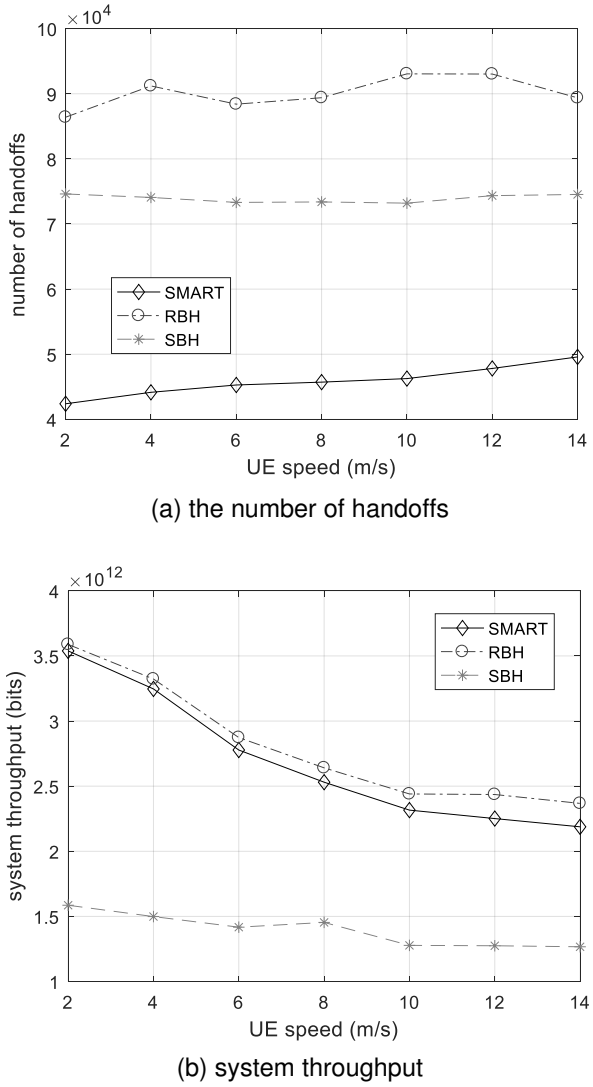
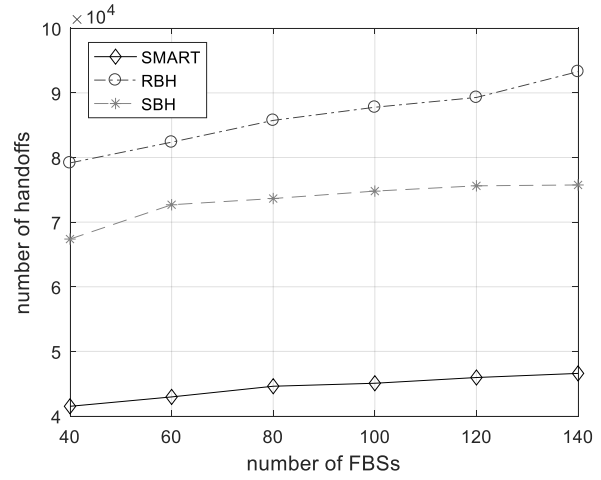


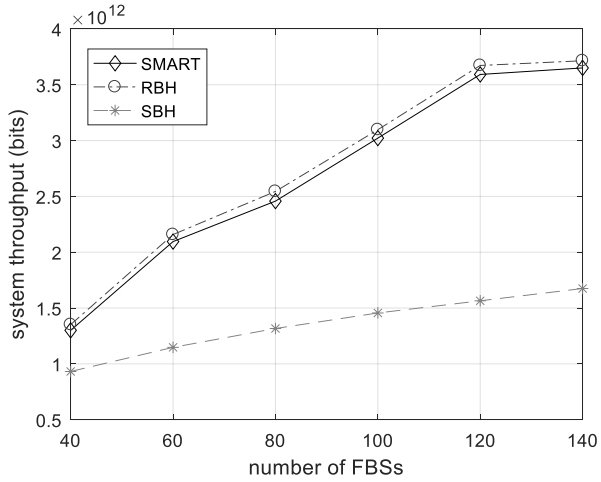
Fig. 6. Relationship between handoff performance and UE speed

and system throughput as a function of the number of FBSs. From Fig.7 (a) we can see that the number of handoffs for SMART is always significantly smaller than that of the other two policies. When the number of FBSs increases from 40 to 140, the number of handoffs for SMART is increased slightly. Fig.7 (b) shows that the throughput of SMART and SBH increases with the number of FBSs due to more available wireless resources.

In Experiment 6, we examine the optimality of SMART policy. SMART-M algorithm cannot achieve the exactly optimum solution due to the relaxation in solving problem (25). Hence, we compare SMART policy with the optimal solution, denoted by SMART-OPT, which is obtained by using integer programming solver, in small scale scenarios. In the experiments, we set the number of BSs to 20, and vary the number of USs from 50 to 200. Other simulation settings are the same with those in Experiment 1. Fig. 8 shows the comparison of SMART-OPT with the other three handoff policies in term of the number of handoffs in 500 seconds. From this figure, we can see that the difference between SMART-OPT and SMART is rather small, which means that SMART policy can reach a near-optimal performance in



(a) the number of handoffs



(b) system throughput

Fig. 7. Handoff performance as a function of the number of FBSs

terms of the number of handoffs. On the other hand, we would like to mention that the computational complexity of SMART is much lower than that of brute force algorithm. We find that the brute force algorithm is at least an order of magnitude lower than the other three policies.

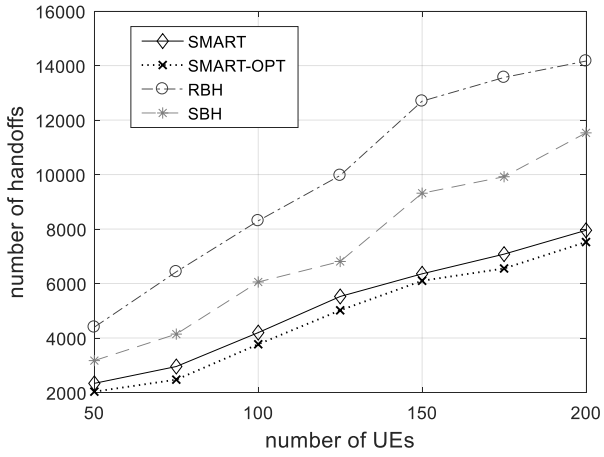


Fig. 8. The comparison of number of handoffs with optimal solution.

8 Conclusions

In this paper, the SMART handoff policy is proposed for mmWave HetNets based on reinforcement learning. In SMART, the handoff trigger conditions are determined by taking into account both mmWave channel characteristics and QoS requirements of UEs. SMART has two BS selection algorithms for different UE density circumstances. SMART-S is for single UE and uses reinforcement-learning for BS selection. SMART-M is for multiple UEs and uses a heuristic for the simultaneous identification of the best target BSs. The computational complexity of SMART is much lower than that of brute force algorithm to calculate the optimal solution. Moreover, as SMART is based on learning, it can be implemented in a distributed manner. Numerical results have shown that the performance of SMART is near the optimal solution. Without sacrificing UE QoS, SMART can reduce the number of handoffs by about 50% when compared with handoff policies without machine learning.

References

[1] B. V. Quang, R. V. Prasad, and I. Niemegeers, "A Survey on Handoffs Lessons for 60 GHz Based," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 1, pp. 64–86, 2012.

[2] G. Gódor, Z. Jakó, Á. Knapp, and S. Imre, "A survey of handover management in LTE-based multi-tier femtocell networks: Requirements, challenges and solutions," *Computer Networks*, vol. 76, pp. 17–41, 2015.

[3] 3GPP TS 36.331, "E-UTRA Radio Resource Control (RRC); Protocol specification (Release 9)," 2016.

[4] Qualcomm Europe, "Range expansion for efficient support of heterogeneous networks," *TSG-RAN WG1*, 2008.

[5] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable Model for Rate in Self-Backhauled Millimeter Wave Cellular Networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2196–2211, 2015.

[6] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, S. T. Rappaport, and E. Erkip, "Millimeter Wave Channel Modeling and Cellular Capacity Evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, 2014.

[7] A. Talukdar, M. Cudak, and A. Ghosh, "Handoff Rates for Millimeterwave 5G Systems," in *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*, 2014, pp. 1–5.

[8] F. Guidolin, I. Pappalardo, A. Zanella, and M. Zorzi, "Context-Aware Handover Policies in HetNets," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 1895–1906, 2016.

[9] A. H. Arani, M. J. Omidi, A. Mehbodniya, and F. Adachi, "A Handoff Algorithm Based on Estimated Load for Dense Green 5G Networks," in *2015 IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–7.

[10] Z. Guohua, P. Legg, and G. Hui, "A network controlled handover mechanism and its optimization in LTE heterogeneous networks," in *IEEE Wireless Communications and Networking Conference, WCNC*, 2013, pp. 1915–1919.

[11] G. Araniti, J. Cosmas, A. Iera, A. Molinaro, A. Orsino, and P. Scopelliti, "Energy Efficient Handover Algorithm For Green Radio Networks," in *2014 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, 2014, pp. 1–6.

[12] H. Leem, J. Kim, D. K. Sung, Y. Yi, and B.-h. Kim, "A Novel Handover Scheme to Support Small-cell Users in a HetNet Environment," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, 2015, pp. 1978–1983.

[13] B. Linh, M. G. Larrode, R. V. Prasad, I. Niemegeers, and A. M. J. Koonen, "Radio-over-Fiber based architecture for seamless wireless indoor communication in the 60 GHz band," *Computer Communications*, vol. 30, no. 18, pp. 3598–3613, 2007.

[14] M. Polese, "Performance Comparison of Dual Connectivity and Hard Handover for LTE-5G Tight Integration in mmWave Cellular Networks," *arXiv preprint arXiv:1607.05425*, vol. 3, 2016. [Online]. Available: http://arxiv.org/abs/1607.04330

[15] M. Mezzavilla, S. Goyal, S. Panwar, S. Rangan, and M. Zorzi, "An MDP Model for Optimal Handover Decisions in mmWave Cellular Networks," in *Networks and Communications (EuCNC), 2016 European Conference on. IEEE*, 2016, pp. 100–105.

[16] M. Wang, A. Dutta, S. Buccapatnam, and M. Chiang, "Smart Exploration in HetNets : Minimizing Total Regret with mmWave," in *IEEE International Conference on Sensing, Communication and Networking*, London, UK, 2016.

[17] M. N. Soorki, M. J. Abdel-rahman, A. Mackenzie, and W. Saad, "Joint Access Point Deployment and Assignment in mmWave Networks with Stochastic User Orientation Joint Access Point Deployment and Assignment in mmWave Networks with Stochastic User Orientation," in *2017 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, no. Paris, 2017, pp. 1–6.

[18] International Telecommunication Union, "Requirements related to technical performance for IMTadvanced radio interfaces," *ITU I.2134*, 2009.

[19] H. Elshaer, M. N. Kulkarni, F. Boccardi, J. G. Andrews, and M. Dohler, "Downlink and Uplink Cell Association With Traditional Macrocells and Millimeter Wave Small Cells," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6244–6258, 2016.

[20] C. Phillips, D. Sticker, and D. Grunwald, "A Survey of Wireless Path Loss Prediction and A Survey of Wireless Path Loss Prediction and Coverage Mapping Methods," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 255–270, 2013.

[21] A. Kumar, D. Manjunath, and J. Kuri, *wireless networking*, 2008.

[22] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*. Oxford, UK: Academic Press, 2007.

[23] M. Giordani, M. Mezzavilla, and M. Zorzi, "Initial Access in 5G mm-Wave Cellular Networks," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 40–47, 2016.

[24] C. N. Barati, S. A. Hosseini, S. Rangan, P. Liu, K. Thanasis, S. S. Panwar, and S. T. Rappaport, "Directional Cell Discovery in Millimeter Wave Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 1–13, 2015.

[25] M. Giordani, M. Mezzavilla, C. N. Barati, S. Rangan, and M. Zorzi, "Comparative Analysis of Initial Access Techniques in 5G mmWave Cellular Networks," in *Information Science and Systems (CISS), 2016 Annual Conference on. IEEE*, 2016, pp. 268–273.

[26] V. Desai, L. Krzymien, A. Soong, A. Alkhateeb, L. Krzymien, P. Sartori, and W. Xiao, "Initial beamforming for mmWave communications," in *Signals, Systems and Computers, 2014 48th Asilomar Conference on. IEEE,*, 2015, pp. 1926–1930.

[27] A. Capone, I. Filippini, and V. Sciancalepore, "Context information for fast cell discovery in mm-wave 5G networks," in *IEEE European Wireless Conference;*, 2015, pp. 1–6.

[28] F. Devoti, I. Filippini, and A. Capone, "Facing the millimeter-wave cell discovery challenge in 5G networks with context-awareness," *IEEE Access*, vol. 4, pp. 8019–8034, 2016.

[29] F. Pantisano, M. Bennis, W. Saad, S. Valentin, and M. Debbah, "Matching with Externalities for Context-Aware User-Cell Association in Small Cell Networks," in *GLOBECOM Workshops (GC Wkshps), 2013 IEEE*, 2013, pp. 4483–4488. [Online]. Available: http://arxiv.org/abs/1307.2763

[30] H. Wang, L. Ding, P. Wu, Z. Pan, N. Liu, and X. You, "QoS-aware load balancing in 3GPP long term evolution multi-cell networks," in *IEEE International Conference on Communications*, 2011, pp. 1–5.

[31] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in Applied Mathematics*, vol. 6, no. 1, pp. 4–22, 1985.

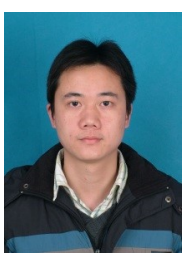
[32] F. P. Auer P, Cesa-Bianchi N, "Finite-time Analysis of the Multiarmed Bandit Problem," *Machine learning*, vol. 47, no. (2-3), pp. 235–256, 2002.

[33] S. H. Low, S. Member, and D. E. Lapsley, "Optimization Flow Control I : Basic Algorithm and Convergence," *IEEE/ACM Transactions on Networking (TON)*, vol. 7, no. 6, pp. 861–874, 1999.

[34] D. P. Bertsekas, *Convex Optimization Theory*. Athena Scientific, 2009.

Yao Sun received the B.S. degree in Mathematical Sciences from the University of Electronic Science and Technology of China (UESTC). He is currently working towards his Ph.D. degree at National Key Laboratory of Science and Technology on Communications, UESTC. His research and study interests include intelligent access control, handoff and resource management in mobile networks based on machine learning and other data analytics.

Gang Feng (M'01, SM'06) received his BEng and MEng degrees in Electronic Engineering from the University of Electronic Science and Technology of China (UESTC), in 1986 and1989, respectively, and the Ph.D. degrees in Information Engineering from The Chinese University of Hong Kong in1998. He joined the School of Electric and Electronic Engineering, Nanyang Technological University in December 2000 as an assistant professor and was promoted as an associate professor in October 2005. At present he is a professor with the National Laboratory of Communications, University of Electronic Science and Technology of China. Dr. Feng has extensive research experience and has published widely in computer networking and wireless networking research. His research interests include resource management in wireless networks, next generation cellular networks, etc. Dr. Feng is a senior member of IEEE.



Shuang Qin received the B.S. degree in Electronic Information Science and Technology, and the Ph.D. degree in Communication and Information System from University of Electronic Science and Technology of China (UESTC), in 2006 and 2012, respectively. He is currently an associate professor with National Key Laboratory of Science and Technology on Communications in UESTC. His research interests include cooperative communication in wireless networks, data transmission in opportunistic networks and green communication in heterogeneous networks.

Ying-Chang Liang (F'11) is a Professor in the University of Electronic Science and Technology of China (UESTC), China, and also a Professor in the University of Sydney, Australia. He was a Principal Scientist and Technical Advisor in the Institute for Infocomm Research (I2R), Singapore. His research interest lies in the general area of wireless networking and communications, with current focus on applying artificial intelligence, big data analytics and machine learning techniques to wireless network design and optimization. Dr Liang was elected a Fellow of the IEEE in December 2010, and was recognized by Thomson Reuters as a Highly Cited Researcher in 2014, 2015 and 2016. He received IEEE ComSocs TAOS Best Paper Award in 2016, IEEE Jack Neubauer Memorial Award in 2014, the First IEEE ComSocs APB Outstanding Paper Award in 2012, and the EURASIP Journal of Wireless Communications and Networking Best Paper Award in 2010. He also received the Institute of Engineers Singapore (IES)s Prestigious Engineering Achievement Award in 2007, and the IEEE Standards Associations Outstanding Contribution Appreciation Award in 2011, for his contributions to the development of IEEE 802.22 standard. Dr Liang is now serving as the Chair of IEEE Communications Society Technical Committee on Cognitive Networks, an Associate Editor of IEEE Transactions on Signal and Information Processing over Network, and an Associate Editor-in-Chief of the World Scientific Journal on Random Matrices: Theory and Applications. He served as Founding Editor-in-Chief of IEEE Journal on Selected Areas in Communications Cognitive Radio Series, and was the key founder of the new journal IEEE Transactions on Cognitive Communications and Networking. He has been an (Associate) Editor of IEEE Transactions on Wireless Communications, IEEE Transactions on Vehicular Technology, and IEEE Signal Processing Magazine. Dr Liang was a Distinguished Lecturer of the IEEE Communications Society and the IEEE Vehicular Technology Society, and has been a member of the Board of Governors of the IEEE Asia-Pacific Wireless Communications Symposium since 2009. He served as Technical Program Committee (TPC) Chair of CROWN08 and DySPAN10, Symposium Chair of ICC12 and Globecom12, General Co-Chair of ICCS10 and ICCS14. He serves as TPC Chair and Executive Co-Chair of Globecom17 to be held in Singapore.

Tak-Shing Peter Yum (F'13) was born in Shanghai. He received primary and secondary school education in Hong Kong. He went to Columbia University and was awarded BS, MS, MPh and PhD degrees in 1974, 1975, 1977 and 1978 respectively. He joined Bell Telephone Laboratories in April 1978 working on switching and signaling systems for 2.5 years. Then, he taught at National Chiao Tung University, Taiwan for 2 years before joining The Chinese University of Hong Kong in 1982. He was appointed chairman of IE Department two times and elected Dean of Engineering for two terms (2004-2010). Since June 1, 2010 he took no-pay leave from CUHK to serve as CTO of ASTRI www.astri.org (Hong Kong Applied Science and Technology Research Institute Company Limited). He is currently a professor at Hunan University.

Professor Yum has published widely in Internet research with contributions to routing, buffer management, deadlock handling, message resequencing and multi-access protocols. He then branched out to work on cellular network, lightwave networks, video distribution networks and 3G networks. His recently research is in the areas of RFID, sensor networks and wireless positioning technologies. He and student Lei Zhu was awarded the Best Paper Award of ACM MSWiM 2009 with paper title, "The Optimization of Framed Aloha based RFID Algorithms." He and another student Xu Chen were awarded the Honorable Mention Award (the first runner-up of the best paper award) with paper title "Cross Entropy Approach For Patrol Route Planning In Dynamic Environments" in IEEE international conference on Intelligence and Security Informatics (ISI), 2010.