Reinforcement Learning based Handoff for Millimeter Wave Heterogeneous Cellular Networks

Yao Sun*,Gang Feng*,Shuang Qin*,Ying-Chang Liang*,Tak-Shing Peter Yum†‡

* National Key Laboratory of Science and Technology on Communications,
University of Electronic Science and Technology of China, Chengdu, China
† College of Computer Science and Electronic Engineering, Hunan University, Changsha, China
‡ Department of Information Engineering, The Chinese University of Hong Kong, Shatin, NT, Hong Kong
Email: fenggang@uestc.edu.cn

***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 if conventional Reference Signal Received Power (RSRP) based handoff mechanism is used. In this paper, we propose a reinforcement learning based handoff policy named LESH to reduce the number of handoffs while maintaining user Quality of Service (QoS) requirements in mmWave HetNets. In LESH, 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, LESH can significantly reduce the number of handoffs when compared with traditional handoff policies.**

I. Introduction

The 5G network needs to dramatically increase network capacity for keeping pace with increasing mobile traffic demand. A simple way to increase network capacity is to allocate more bandwidth. 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 occurs more frequently in mmWave HetNets. It was shown in [2] 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.

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.

There are relatively few papers on handoff in mmWave HetNets. The authors of [4] proposed the Extended Cell (EC) concept in RoF architecture to increase overlapping areas and thus decrease handoff UE outage probability. Similarly, to the support of high UE mobility in outdoor environment, the Moving Extended Cell [5] and Moving Extended N-Cells [6] concepts are proposed. Focusing on the optimization of handoff mechanisms, the authors of [7] solved the BS selection problem by Markov Decision Process (MDP). 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 be applied to dense deployment HetNets. The authors of [8] develop an online learning-based approach to solve single UE network selection problem in heterogeneous wireless networks which contains 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 long-term throughput of the UE.

In this paper, we propose the Learning based Smart Handoff (LESH) policy for mmWave HetNets. Our design objective is to reduce the number of unnecessary handoffs while guaranteeing the QoS of UEs. LESH 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: LESH-S and LESH-M for different UE density circumstances. LESH-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. LESH-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.

II. System Model

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 BS (MBS). Let λ be the ratio of FBSs using mmWave frequency, M_m be the set of the mmWave FBS (denoted as mm-FBS), and M_t be the set of the traditional FBS (denoted as Tr-FBS).

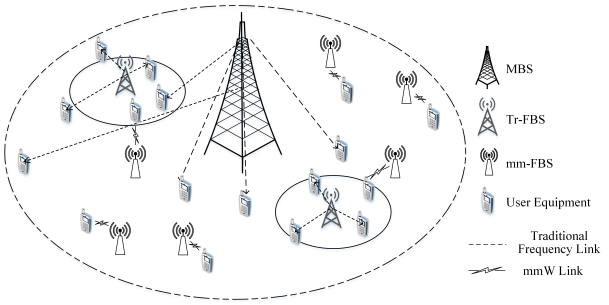


Fig. 1. The system model of mmWave HetNets.

We assume that the channel of mm-FBS is based on LOS-NLOS models [9], meaning that the channel condition between UE and mm-FBS can alternate between the two well-defined states, Line-of-Sight (LOS) and Non-Line-of-Sight (NLOS). Similar to that in [10], [11], 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, α and η are the least square fits of floating intercept and slope over the measured distances (30 to 200 m), and σ^2 is the lognormal shadowing variance. The values of α, η and θ are different for LOS and NLOS states [10]. 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 $SNR_n^j = \frac{P_j \psi L(d)^{-1}}{\sigma^2}$, where P_j is the transmit power of mm-FBS j, σ^2 is the noise power and ψ is the antenna gain that can be calculated according to [10].

For the traditional band, co-channel interference needs to be considered due to shared bandwidth deployment. 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

$$r_n^j = \begin{cases} \frac{B_m}{U_j} log_2(1 + SNR_n^j), & j \in \mathcal{M}_m \\ \frac{B_t}{U_j} log_2(1 + SINR_n^j), & j \in \{\mathcal{M}_t \cup MBS\} \end{cases}, \quad (2)$$

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.

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

$$\exists t_0 \in [t - \tau_n, t], s.t. r_n^j(t_0) \geq \gamma_n^{min}. \quad (3)$$

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.

III. Framework of LESH Handoff Policy

3GPP defines six handoff events for cellular networks [3] with Event A2 and Event A3 being the most two common events in HetNets. Our proposed LESH handoff mechanism is focused on these two handoff events, and the other handoff events decisions remain the same as those in 3GPP.

A. Handoff Trigger Conditions

Event A2 occurs when the serving BS becomes worse than a threshold [3], or the serving BS cannot fulfill the minimum UE QoS requirement. Thus, in LESH, the trigger condition can be expressed as

$$\forall t_0 \in [t - \tau_n, t], r_n^i(t_0) < \gamma_n^{min}, \quad (4)$$

where τ_n and γ_n^{min} are UE service type parameters. This change can avoid many unnecessary handoffs. Once inequality (4) 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 amount better than the serving BS [3]. In this event, the UE switches to a better BS which can improve his QoS although current serving BS can fulfill the minimum QoS requirement. Thus, LESH uses the following three trigger conditions

$$\exists t_0 \in [t - \tau_n, t], s.t. r_n^j(t_0) \geq \gamma_n^{min}, \quad (5\text{-}1)$$

$$r_n^k(t) \geq r_n^j(t) + \mathit{offset}, \quad (5\text{-}2)$$

$$\gamma_n^{max} - \gamma_n^{min} > \epsilon. \quad (5\text{-}3)$$

Condition (5-1) states that the current serving BS can fulfill the minimum UE QoS requirement. Condition (5-2) constraints that the transmission rate of the target BS k is at least *offset* higher than that of the serving BS j. Condition (5-3) indicates that the difference of transmission rate between maximum and minimum threshold is greater than ϵ in QoS requirement.

B. BS Selection

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

IV. LESH-S Algorithm for Single Target BS Selection

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,

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

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.

A. 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.

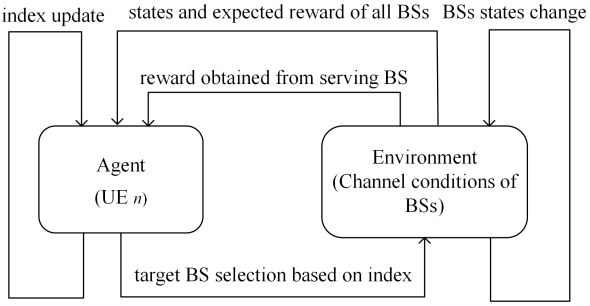


Fig. 2. Reinforcement learning based BS selection framework

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

$$R_n^k(t) = \int_t^{t_n^k} r_n^k(t)dt. \tag{6}$$

***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 (6).*

Proof: Let t_n^k in (6) 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

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

The objective of the above reinforcement learning model is to find the optimal policy $\pi^* = \arg\max_\pi \mathbb{E}_\pi \left[\sum_{\Phi(t_n^k)=1}^{K} R_n^k(t)\right]$, 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 $\min H_n$. ■

B. Expected Reward Estimation

As t_n^k and $r_n^k(t)$ in (6) 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 from 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

$$\check{R}_{C_n}^k(0) = 0, \tag{7-1}$$

$$\check{R}_{C_n}^k(T_{C_n}^k + 1) = \frac{T_{C_n}^k \check{R}_{C_n}^k(T_{C_n}^k) + \check{R}_n^k(t)}{T_{C_n}^k + 1}, \tag{7-2}$$

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 (7-2). Thus, the expected reward can be estimated as

$$\mathbb{E}[R_n^k(t)] = \check{R}_{C_n}^k(T_{C_n}^k), \text{ for } n \in C_n \tag{8}$$

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.

C. BS Selection Algorithm

We cannot always select the BS with the highest reward since a well-known dilemma exploration vs. exploitation exists. 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). Based on UCB algorithm, we propose a BS selection policy T when UE n triggers Event A2 handoffs. We set index of BS j for UE n as $\mathbb{E}[R_n^k(t)] + \ell\sqrt{\frac{2\ln H_n}{T_{C_n}^k}}$, 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

$$k^* = \arg\max_k \left(\mathbb{E}[R_n^k(t)] + \ell\sqrt{\frac{2\ln H_n}{T_{C_n}^k}}\right). \tag{9}$$

We summarize LESH-S BS selection algorithm as follows:

V. LESH-M Algorithm for Multiple Target BS Selection

The BS selection algorithm discussed in Section IV 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 LESH-M algorithm for optimal multi-BS selection.

Algorithm 1 : LESH-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 $T^k_{C_n}$, H_n, $\check{R}^k_{C_n}(T^k_{C_n})$ 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
    $\check{R}^k_{C_n}(T^k_{C_n}+1) \leftarrow \frac{T^k_{C_n}\check{R}^k_{C_n}(T^k_{C_n})+\check{R}^k_n(t)}{T^k_{C_n}+1}$
    $k^* = \arg\max_k \left(\mathbb{E}[R^k_n(t)] + \ell\sqrt{\frac{2\ln H_n}{T^k_{C_n}}}\right)$
    $T^k_{C_n} \leftarrow T^k_{C_n}+1$,$H_n \leftarrow H_n+1$
  else
    switch to the unique target BS k^*
  end if
end while

A. 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. 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 the period. Here, the objective function Y is again chosen as the volume of transmitted data before the next handoff occurs for these N UEs. The problem is formulated as

$$\max\ Y = \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{A}_i} x_{ij}\mathbb{E}[R^k_n(t)] \tag{10}$$

$$\text{s.t.} \sum_{i\in\mathcal{N}} x_{ij} \le N_j, \forall j \in \cup_{i\in\mathcal{N}}\mathcal{A}_i, \tag{10-1}$$

$$\sum_{j\in\mathcal{A}_i} x_{ij} = 1, \forall i \in \mathcal{N} \tag{10-2}$$

$$x_{ij} \in \{0,1\}, \forall i \in \mathcal{N}, \forall j \in \cup_{i\in\mathcal{N}}\mathcal{A}_i, \tag{10-3}$$

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 (10-1) ensures that the number of UEs which switch to the same BS does not exceed the current BS connection capacity. Constraints (10-2) and (10-3) 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.

B. BS Selection Algorithm

The problem stated in (10) is a special case of a well-known NP-hard problem "Generalized Assignment Problem (GAP)", we propose the following efficient heuristics. We first relax binary variables x_{ij} in constraints (10-3) to be continuous variables in $[0,1]$. We then exploit Lagrange dual decomposition method [14] to solve this optimization problem.

After relaxing x_{ij}, problem (10) becomes a linear problem with Lagrange function $L(\boldsymbol{x},\boldsymbol{\mu}) = \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{A}_i} x_{ij}\mathbb{E}[R^k_n(t)] - \sum_{j\in\mathcal{A}}\mu_j(\sum_{i\in\mathcal{N}} x_{ij} - N_j)$, where μ_j is Lagrange multiplier. For a fixed vector $\boldsymbol{\mu}$, Lagrange dual function can be expressed as

$$g(\boldsymbol{\mu}) = \sup_{\boldsymbol{x}}\ L(\boldsymbol{x},\boldsymbol{\mu}) \tag{11}$$

$$\text{s.t.} \sum_{j\in\mathcal{A}_i} x_{ij} = 1, \forall i \in \mathcal{N}, \tag{11-1}$$

$$0 \le x_{ij} \le 1, \forall i \in \mathcal{N}, \forall j \in \mathcal{A}, \tag{11-2}$$

and the dual problem is $\min_{\boldsymbol{\mu}} g(\boldsymbol{\mu})$. Rewriting function $g(\boldsymbol{\mu})$ yields $g(\boldsymbol{\mu}) = \sup_{\boldsymbol{x}}\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{A}_i} x_{ij}(\mathbb{E}[R^k_n(t)]) - \mu_j) + \sum_{j\in\mathcal{A}}\mu_j N_j$. Since it dose not include the cross-term of x_{ij}, we can exchange the computation order as:

$$g(\boldsymbol{\mu}) = \sum_{i\in\mathcal{N}}\sup_{x_{ij},j\in\mathcal{A}_i}\sum_{j\in\mathcal{A}_i} x_{ij}(\mathbb{E}[R^k_n(t)] - \mu_j) + \sum_{j\in\mathcal{A}}\mu_j N_j.$$

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

$$g_i(\boldsymbol{\mu}) = \sup_{x_{ij},j\in\mathcal{A}_i}\sum_{j\in\mathcal{A}_i} x_{ij}(\mathbb{E}[R^k_n(t)]) - \mu_j) \tag{12}$$

$$\text{s.t.} \sum_{j\in\mathcal{A}_i} x_{ij} = 1, \tag{12-1}$$

$$0 \le x_{ij} \le 1. \forall j \in \mathcal{A}_i. \tag{12-2}$$

Since we want to find a binary solution of x_{ij}, for a fixed vector $\boldsymbol{\mu}$, problem (12) is described as: for UE i, we choose a BS j^* from set $\mathcal{A}_i$ to maximize the value of $\mathbb{E}[R^{k^*}_n(t)] - \mu_{j^*}$. Therefore, when $\boldsymbol{\mu}$ is fixed, problem (11) 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 $\mu_j \ge 0$,

$$\mu_j(k+1) = \left[\mu_j(k) - \delta(k)(N_j - \sum_{i\in\mathcal{N}} x_{ij})\right]^+, \forall j \in \mathcal{A}, \tag{13}$$

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

$$\delta(k) = \frac{g(\boldsymbol{\mu}_k) - g_k}{\|\boldsymbol{h}_k\|^2}, \forall k \ge 0, \tag{14}$$

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

$$g_k = \min_{1\le j\le k} g(\boldsymbol{\mu}_k) - \varepsilon_k, \tag{15}$$

and ε_k is updated according to

$$\varepsilon_{k+1} = \begin{cases} \rho\varepsilon_k & \text{if } g(\boldsymbol{\mu}_{k+1}) \le g_k \\ \max\{\beta\varepsilon_k, \varepsilon\} & \text{otherwise} \end{cases}, \tag{16}$$

where ε, β and ρ are fixed positive constant with $\beta < 1$ and $\rho \ge 1$ [15]. For linear programs, strong duality holds. Therefore, the minimum value of $g(\boldsymbol{\mu})$ is equal to the maximum value of the original problem. Similar to that in Section IV, $\mathbb{E}[R^k_n(t)]$ is updated once the next handoff occurs according to (7) and (8). Note that, the reinforcement-learning process in

Section IV can improve the accuracy of the value of $\mathbb{E}[R_n^k(t)]$ thus the solution of this optimization problem. We summarize the LESH-M algorithm in Algorithm 2.

Algorithm 2 : Joint optimal LESH-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 $\check{R}_{C_i}^j(T_{C_i}^j)$ 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** $\boldsymbol{x}^k \neq \boldsymbol{x}^{k-1}$ **do**
6: $k \leftarrow k+1$
7: **for** each UE $i \in \mathcal{N}$ **do**
8: solve problem (12)
9: **end for** (obtain $\boldsymbol{x}^k$)
10: update $\boldsymbol{\mu}^k$ according to (13)
11: **end while**
12: $\boldsymbol{x}^* \leftarrow \boldsymbol{x}^k$

VI. Numerical Results

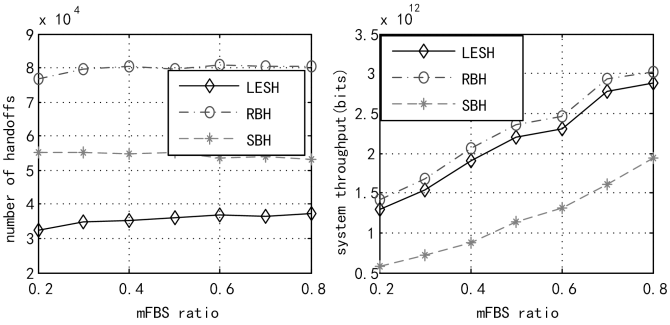
We now compare the performance of LESH with two conventional handoff policies as follows. (1) Rate-based handoff (RBH). RBH has similar trigger conditions as those in 3GPP. When choosing target BSs for handoffs, the ones with maximum transmission data rates are chosen (instead of maximum RSRP in 3GPP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

A. Simulation Settings

We consider a two-tier HetNet which consists of an MBS and varying number of mm-FBSs, Tr-FBSs and UEs. Both mm-FBSs and Tr-FBSs are randomly distributed. The transmit power of MBS, mm-FBS and Tr-FBS are set to 46dBm, 30dBm and 20dBm, respectively. Both the numbers and regions of blockages in mm-FBS are randomly generated. Similar to [11], 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). We use $L(d) = 34 + 40\log(d)$ and $L(d) = 37 + 30\log(d)$ to model the path loss for the MBS and Tr-FBSs respectively [16]. The bandwidth allocated to MBS/Tr-FBSs and mm-FBSs are 20MHz and 500MHz respectively. The noise power is set to -101dBm and -77dBm for traditional and mmWave band respectively [10]. We assume that UEs are randomly distributed in the area and move to a random direction at a random speed.

B. 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 LESH is 7.8×10^4, $5.5 \times 10^4$and 3.2×10^4, respectively. These numbers show that LESH can reduce handoffs to 41% and 58% when compared with RBH and SBH respectively. For $\lambda = 0.8$, the reduction percentages are 46% and 68%. 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 LESH and RBH is relatively small (3% for $\lambda = 0.8$, 6% for $\lambda = 0.2$), implying that significant handoff performance gain can be accomplished with a small compromise on throughput.

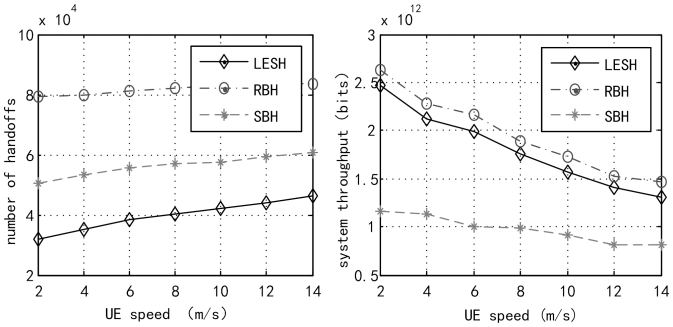


(a) the number of handoffs (b) system throughout

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

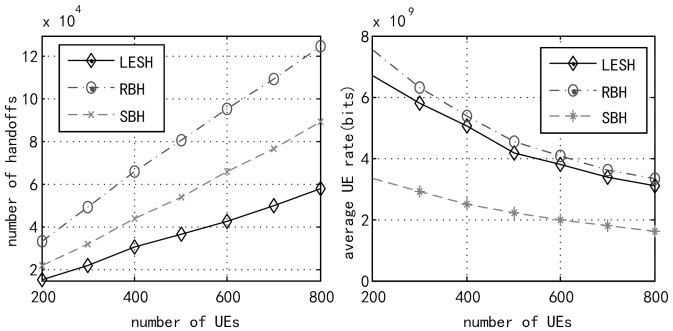
In Experiment 2, we examine the effect of UE movement speed at $\lambda = 0.5$ with parameters the same as the Experiment 1. Fig.4 shows the number of handoffs and system throughput for the three handoff policies as a function of the mean UE movement speed. We see that from fast walking speed of 2 m/s (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 LESH remains. As expected, Fig.4 (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 3, we compare the performance of handoff policies by the number of UEs at =0.5. Fig.5 (a) shows that for a wide range of UE densities, the number of handoffs for LESH is only about 50% of RBH and 70% of SBH. We also see that the difference of the number of handoffs between



(a) the number of handoffs (b) system throughout

Fig. 4. Relationship between handoff performance and UE speed.



(a) the number of handoffs (b) average UE transmission rate

Fig. 5. Handoff performance as a function of the number of UEs.

LESH and the other two handoff policies increases with the number of UEs. This implies that LESH-M algorithm has even better performance in dense UE conditions. Fig.5 (b) shows that the average UE transmission rate for the three handoff policies decreases with the number of UEs due to wireless resource limitations.

VII. Conclusions

In this paper, the smart handoff policy LESH is proposed for mmWave HetNets based on reinforcement learning. In LESH, the handoff trigger conditions are determined by taking into account both mmWave channel characteristics and QoS requirements of UEs. LESH has two BS selection algorithms for different UE density conditions. LESH-S is for single UE and uses reinforcement-learning for BS selection. LESH-M is for multiple UEs and uses a heuristic for the simultaneous identification of the best target BSs. Numerical results show that, without sacrificing UE QoS, LESH 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] A. Talukdar, M. Cudak, and A. Ghosh, "Handoff Rates for Millimeter-wave 5G Systems," in *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*, 2014, pp. 1–5.

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

[4] 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.

[5] K. Tsagkaris, N. D. Tselikas, and N. Pleros, "A Handover Scheme Based on Moving Extended Cells for 60 GHz Radio-Over-Fiber Networks," in *IEEE International Conference on Communications*, 2009, pp. 1–5.

[6] N. D. Tselikas and A. C. Boucouvalas, "Evaluating Dominant Handoff Schemes for Vehicular Radio-over-Fiber Networks @ 60GHz," in *2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)*, 2014, pp. 83–88.

[7] M. Mezzavilla, S. Goyal, S. Panwar, S. Rangan, and M. Zorzi, "An MDP Model for Optimal Handover Decisions in mmWave Cellular Networks," *arXiv preprint arXiv:1507.00387,*, 2015.

[8] 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*, 2016.

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

[10] S. Singh, M. N. Kulkarni, S. Member, 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.

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

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

[13] F. P. Auer P, Cesa-Bianchi N, "Finite-time Analysis of the Multiarmed Bandit Problem," *Machine learning*, pp. 235–256, 2002.

[14] 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.

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

[16] Q. Ye, B. Rong, Y. Chen, M. Al-shalash, C. Caramanis, and J. G. Andrews, "User Association for Load Balancing in Heterogeneous Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, 2013.