Optimal Handover Policy for mmWave Cellular Networks: A Multi-Armed Bandit Approach

Li Sun, Jing Hou, and Tao Shu
Department of Computer Science and Software Engineering
Auburn University, Auburn, AL, 36849, USA
{lzs0070, jzh0141, tshu}@auburn.edu

***Abstract*—Millimeter wave (mmWave) is a promising technology in 5G communication due to its abundant bandwidth resource. However, its severe path attenuation and vulnerability to line-of-sight (LOS) blockage result in much more unpredictable outages than traditional technologies. This special propagation property raises a significant challenge to the mobility management in mmWave cellular networks. In particular, conventional handover policies purely rely on the measurement of signal strength. If being applied directly in mmWave cellular networks, they would cause a large number of unnecessary handovers due to the frequent short-term LOS blockage by obstacles, imposing high signaling and energy overhead on the network. In this paper, we propose a novel handover mechanism to reduce unnecessary handovers in a mmWave cellular network by carefully deciding the next base station (BS) a user should handover to, so that the new user-BS connection after the handover can last as long as possible. Clearly, making such an optimal decision requires some knowledge on the users' post-handover mobility trajectory and LOS blockage, whose realization cannot be assumed at the moment of handover. The proposed handover mechanism addresses this challenge by exploiting the *empirical distribution* of users' post-handover trajectory and LOS blockage, learned online through a multi-armed bandit framework, with the intention to maximize the expectation of the user-BS connection time after each handover. The results of numerical experiments demonstrate that the proposed handover policy outperforms existing counterparts on reducing handovers, especially in the scenarios where users' mobility follows regular patterns.**

***Index Terms*—wireless communication, millimeter wave, handover, online learning, multi-armed bandit**

I. Introduction

Millimeter wave (mmWave) is one of the fundamental technologies in the upcoming 5G cellular networks. Because of its 10-to-100s GHz level frequency, mmWave can provide abundant bandwidth for wireless service through the line-of-sight (LOS) path. However, a big challenge for mmWave to be utilized in practical cellular networks is that mmWave communication heavily relies on the LOS path, but this path could be easily blocked by obstacles (e.g., tree-tops, pedestrians, and buildings) with the movement of the user. Due to its short wavelength, once the LOS is blocked, the mmWave signal will not be able to penetrate through or circumvent around the obstacle, leading to sudden significant drop of the received signal (a.k.a. outage), which urges the user equipment (UE) to handover to another base station (BS) in order to maintain the connection. As such, it has been shown in the literature that the handover frequency in mmWave cellular networks is much higher than that in current 4G systems [1]. Therefore, efficient mobility and handover management is an inherent challenge that needs to be addressed in mmWave cellular networks.

The existing studies on mmWave handover management are mainly focused on two directions: increasing handover success rate and reducing unnecessary handovers. Multi-connectivity is a solution to provide reliable service and reduce handovers [2]. In [3], [4], the authors proposed multi-connectivity protocols and specified evaluation methods for handover in mmWave cellular systems. In order to handle the unexpected LOS link blockage, [5] introduced a caching scheme, which stores extra incoming data frames when high-throughput links are available, and uses the cached content to smooth the transition when handover happens. In [6], [7], the authors modeled the BS selection process as a Markov Decision Process (MDP) by taking account of dynamic channel condition and user mobility to improve the network capacity. Moreover, machine learning provides another promising tool to improve handover decision [8]–[10]. The authors in [8] introduced a reinforcement-learning based handover policy to reduce the number of handovers in HetNet. Furthermore, deep learning was also successfully utilized to implement proactive handover in mmWave band to reduce handover failure rate [9], [10].

Although handovers are frequent in mmWave systems, it has been shown that about 61% handovers are unnecessary or could have been avoided if the UE had made a better choice regarding which BS it should handover to [11]. Reducing unnecessary handovers not only avoids high signaling overhead in the network but also makes an ongoing communication connection smoother. Conventional handover mechanisms are based on measurement of signal strength, and do not perform well in mmWave networks since it may cause "short-sighted" handover decision. For example, a BS with the highest signal strength would be chosen by conventional solutions as the handover target even if the LOS link associated with it will be lost in the next second after the handover. Instead, if another BS that has a lower signal strength but a longer unobstructed time for its LOS path were selected, a redundant handover could have been avoided. Therefore, an optimal handover policy should take into account not only the current instantaneous state of the candidate BSs, but also the future change of state, so as to reach a "far-sighted" handover decision.

Clearly, making such an optimal decision requires some knowledge on the user's post-handover mobility trajectory

978-1-7281-0962-6/19/$31.00 ©2019 IEEE

and LOS blockage, whose realization cannot be assumed at the moment of handover. One straightforward way to address this problem is to predict the user's post-handover trajectory based on his trajectory before the handover. However, this solution requires exact location information of the user (i.e., geo-coordinates of user's location), which is not always available/practical in reality.

In this paper, we propose a handover mechanism that addresses this challenge by exploiting the *empirical distribution* of users' post-handover trajectory and LOS blockage, learned *online* through a multi-armed bandit (MAB) framework, with the intention to maximize the expectation of the user-BS connection time after each handover. In our mechanism, learning happens in the hindsight: each UE will report its reward (definition will be clear shortly) in connecting with a BS in previous handover when it needs to handover again. The accumulative rewards associated with a BS, collected from all users that have connected with that BS when they handover in the same area, serves as a comprehensive indicator of the reward a future UE can expect to receive if it choose to connect with that BS in a handover happening in that area. The MAB construct ensures that the above learning process will converge, and a user can maximize its expected reward by selecting the right BS according to the proposed algorithm. In contrast to the aforementioned trajectory prediction method, an advantage of our mechanism is that it does not require users' exact location information. Instead, user's coarse-grained location information, e.g., the area where the handover happens, is used as index in our algorithm to collect rewards. In practice, these coarse-grained information could be represented as the vector of received signal strength (RSS) from surrounding BSs, and hence is considered practical according to 3GPP [12].

In the literature, our work is most related to the SMART scheme [8], which also uses a reinforcement learning framework to guide BS selection in handover. The main difference between our work and SMART is that our MAB learning model considers the area where a handover happens to better characterize the distribution for post-handover user's trajectory and LOS blockage, while SMART is completely independent from user's location. Our performance evaluation simulates SMART as a counterpart scheme and shows that the proposed mechanism outperforms SMART significantly.

The rest of this paper is organized as follows. In Section II, we describe our system model. In Section III, we propose the online learning handover policy in detail. In Section IV, we present the MAB-based BS selection algorithm. In Section V, we compare the performance of the online learning handover policy with some existing handover policies. Finally, we conclude the paper in Section VI.

II. System Model

A. Network Scenario

Consider a cellular network N consisting of a set of mmWave small cell base stations (SBSs), denoted as B. These SBSs are randomly distributed in the network to provide high throughput by LOS links to UEs in small cells. Actually, SBSs and macro base stations (MBSs) always coexist to provide reliable wireless service. Since MBS can provide larger coverage and is flexible to obstacles because of the conventional sub-6 GHz frequencies, it is used for transmission of controlling signals and acts as a substitution whenever no LOS link is available. A centralized controller (CC) takes charge of handover in this network.

In order to investigate the characteristics of handover in mmWave domain, we only focus on the interaction between SBS and UE in this paper. The switch between SBS and MBS, as well as the interaction between MBS and UE, are not within the scope of our discussion.

B. Propagation Model

In this paper, we assume that the channel of mmWave SBSs is described by 3GPP Standard probabilistic LOS model. According to [8], [13], the statistic path loss model is

$$PL(d)[dB] = \alpha + 10\beta \log_{10}(d) + \xi, \xi \sim N(0, \sigma^2), \qquad (1)$$

where d is the distance in meters, α and β are the least square fits of floating intercept and slop over the measured distances, and σ^2 is the lognormal shadowing variance. Since inter-user interference can be ignored in mmWave band, we only model the SNR of the signal received by the UE n from the SBS $k \in B$ as [8]

$$SNR_n^k = \frac{P_k \times G \times PL(d)^{-1}}{P_n}, \qquad (2)$$

where P_k is the transmit power of SBS k, P_n is the noise power and G is the antenna gain. The antenna gain in mmWave communication highly depends on the direction of beams formed by transmitter and receiver. Since we assume that SBS is equipped with directional antennas with a sectorized gain pattern while UE is equipped with ominidirectional antennas, the UE receives a signal with the channel gain G which is the function of the angle θ between the UE and the SBS. According to [14], this channel gain is calculated as

$$G(\theta) = \begin{cases} G_{max}, & \text{if } |\theta| \le \theta_s \\ G_{min}, & \text{otherwise}, \end{cases} \qquad (3)$$

where G_{max} is the main lobe gain, G_{min} is the side lobe gain, and θ_s is the main lobe width of the SBS. We assume that perfect beam tracking technique can be used to maintain mmWave link [8]. Therefore, the UE could always be in the main lobe and have main lobe gain.

A SBS may serve multiple UEs simultaneously. We assume the served UEs equally share the bandwidth provided by the SBS. The transmission rate for a UE n who is served by the SBS k is calculated as follows:

$$h_n^k = \frac{B_w}{U_k} \log_2(1 + SNR_n^k), \qquad (4)$$

where B_w is the bandwidth provided, and U_k is the total number of UEs currently served by the SBS k. Additionally, since the number of beams formed by SBS is limited, we assume U_{max} as the maximum number of UEs that can be supported by SBS simultaneously.

C. Blockage Model

Since mmWave signal is highly vulnerable to obstacles, we assume that the transmission rate of LOS link will drop to zero immediately when the link is blocked. In this paper, an obstacle is modeled as a circle with a fixed center and a certain radius. Given a LOS link and a set of obstacles, the link is blocked whenever there is an obstacle to which the distance from the link is less than its radius.

III. Online Learning Handover Policy

Within the six handover events defined by 3GPP standard, we focus on the BS selection for Event A2, which is common but challenging in mmWave band. Note that this online learning handover policy is also suitable for Event A1. At a high level, the framework of the proposed MAB-based online learning handover mechanism is illustrated in Fig. 1, and elaborated in the following.

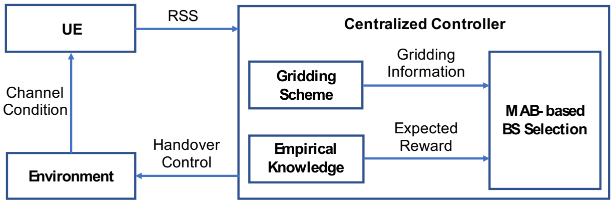


Fig. 1. Framework of online learning handover policy

A. Gridding Scheme

During a handover, instead of picking the BS that has the highest instantaneous received signal strength (RSS), we prefer a BS that has the longest connection time of its LOS path subject to a minimum RSS requirement. The unobstructed time of LOS path is determined by users' post-handover trajectory and the distribution of obstacles around that trajectory. These two key factors are closely related to the area where the handover event occurs. In this section, we introduce a gridding scheme, which partitions the network into a set of grids, each covering a small area of the network. This setting allows handovers to be differentiated by the areas where they happen.

Specifically, in an ultra-densified 5G network, it is common that multiple mmWave small cells overlap. Therefore, a UE at any location is likely to receive from multiple nearby mmWave SBSs, with different RSSs. The id of these SBSs, along with the RSSs from these SBSs, constitute a vector that can be used to identify the instantaneous location of the UE. In contrast to those localization algorithms, our goal here is not to calculate the location of the UE, but to use the above signal-space vector as an index to label the small geographic area (i.e., the grid) that the UE resides in. In particular, two UEs m and n are considered to be in the same grid if they can receive from the same subset of SBSs and the proximity of their RSS vectors $\|v_m - v_n\| \leq \delta$, where $\|\cdot\|$ denotes the Euclidean distance, and δ is a system parameter that controls the size of the grid. In this way, the region of the entire network N can be partitioned into a group of M grids and each grid, denoted by g_i, satisfies $\bigcup_{i=1}^{M} g_i = N$ and $\bigcap_{i=1}^{M} g_i = \emptyset$. For each grid, we choose a representative RSS vector as the indicator of this grid. This vector could be measured manually through a site survey, or be calculated automatically according to some clustering algorithm based on UEs' RSS report (the log file at SBSs contains such information). We assume that CC has full knowledge on the above grid information. Whenever a handover event occurs, CC collects the UE's instantaneous RSS vector, based on which it can tell which grid this handover resides in. Note that in the proposed scheme UE's exact location information is not required. Instead, it uses UE's RSS vector as an index to label the coarse-grained area (i.e., the grid) of the UE, and hence is practical according to 3GPP [12].

B. BS Selection based on Empirical Distribution of Post-handover Trajectory

CC maintains M independent MAB processes, one for each grid. The MAB process for grid g_i is responsible for deciding the optimal SBS a UE should switch to when it has to handover in that grid. In particular, suppose a UE in grid g_i is able to receive from n_i SBSs, denoted by set B_i. Then the MAB process for grid g_i has n_i arms, each representing a distinct SBS in B_i. The MAB process maintains an accumulated reward for each arm. As will be clear shortly, for arm k, where $1 \leq k \leq n_i$, this accumulated reward is calculated by taking into account the rewards received by all historical UEs who switched to SBS k in past handovers that happened in grid g_i, so it reflects the mean reward a future UE is expected to receive if it switches to SBS k after a handover in grid g_i.

An incoming UE that handovers in the grid will switch to the particular SBS whose representing arm presents the highest accumulated reward among all arms. Our MAB construct ensures that when the algorithm converges, the regret between the SBS selected by the algorithm and the SBS selected optimally in the hindsight will be minimized. The actual reward received by this UE, which reflects the actual unobstructed LOS connection time between the previous handover and the next handover, will be computed and reported to the MAB process at CC to update the accumulated reward of the relevant arm when the next handover is due. Clearly, the computation of the accumulated reward for each arm in grid g_i is based on all historical realizations of UEs' post-handover trajectories for handovers in g_i, and hence it is an expectation over the empirical distribution of UE's post-handover trajectory.

C. Handover Trigger Condition

To guarantee the quality of service, the handover trigger condition for a UE n associated with SBS k is described as

$$h_n^k < h_{min} - Hys, \quad (5)$$

where h_{min} is the minimum transmission rate required for a certain service level, and Hys is a hysteresis parameter for avoiding frequent handover. Although how to select a proper value for Hys is an interesting issue, it is not the key point of this paper. For simplicity, we set Hys to be zero. Note that specific value of Hys does not influence the proposed policy.

IV. MAB-based Algorithm for BS Selection

Since our goal is to find the SBS which can bring the longest unobstructed time of LOS path for each handover without prior knowledge, we model the SBS selection in each grid as a MAB problem, which is to identify which arm to pull in order to get maximum reward after a given set of trials [15].

A. Stateless Multi-armed Bandit Model

The representing RSS vector of grid g_i indicates the candidate SBS set B_i for the UEs who reside in the grid. Let h_i^k be the transmission rate received by a UE from SBS k in grid g_i, then B_i is specified as

$$B_i = \{k \mid h_i^k \geq h_{min}, k \in B\}. \tag{6}$$

After CC chooses a SBS $k \in B_i$ in a handover which happens in grid g_i at time τ, the UE will be served by SBS k until it needs another handover, suppose at time τ_i^k. Then the UE receives an instantaneous reward associated with SBS k in grid g_i, denoted as $r_{i,\tau}^k = \tau_i^k - \tau$. Since τ_i^k is an unknown random variable which is determined by the realization of user's post-handover trajectory, the reward $r_{i,\tau}^k$ is also an i.i.d. random variable. As there are no explicit states of SBS as prior knowledge during handover, the SBS selection in grid g_i is formulated by a *stateless* MAB model [15] $\mathcal{M}_i = \{B_i, \mu_{k,t}^i\}$, where $k \in B_i$, and $\mu_{i,t}^k$ is the expected reward of SBS k in grid g_i at round t (means the t-th playing of the model).

Denote $a_{i,t}$ to be the SBS actually selected by CC following a certain policy, in grid g_i at round t. The regret of this policy up to round T, which is defined as the accumulated difference between the reward obtained following this policy and the optimal reward could be obtained with full knowledge, is

$$R_{i,T} = \max_{k \in B_i} \mathbb{E}\left[\sum_{t=1}^{T} r_{i,t}^k\right] - \mathbb{E}\left[\sum_{t=1}^{T} r_{i,t}^{a_{i,t}}\right]. \tag{7}$$

Based on the model $\mathcal{M}_i$, the handover decision problem in grid g_i with the aim to choose the SBS which brings the longest unobstructed LOS connection time, is equivalent to find the optimal policy for the corresponding stateless MAB problem that minimizes the regret.

B. Estimation of Expected Reward

If full knowledge about the distribution of each SBS's reward is known, the optimal policy is to choose the optimal SBS $k^* = \arg\max_{k \in B_i} \mu_{i,t}^k$ for handover in grid g_i all the time. Unfortunately, this assumption does not hold. Therefore, the expected reward of SBS can only be estimated based on historical observations [8]. Denote T_i^k and $\bar{r}_i^k(T_i^k)$, as the number of times that SBS k is chosen and the sample mean of reward of SBS k in grid g_i, respectively. These two metrics are updated by an observation of reward $r_{i,\tau}^k$ as follows:

$$\bar{r}_i^k(T_i^k + 1) = \frac{T_i^k \times \bar{r}_i^k(T_i^k) + r_{i,\tau}^k}{T_i^k + 1}, \tag{8}$$

$$T_i^k := T_i^k + 1. \tag{9}$$

Initially, we set $T_i^k = 0$ and $\bar{r}_i^k(0) = 0$. We use this sample mean value $\bar{r}_i^k(T_i^k)$ as the estimation of the expected reward of SBS k in grid g_i. Each instantaneous reward obtained by any UE is used to update the corresponding mean reward of its serving SBS.

C. Exploration and Exploitation

How to trade off exploration and exploitation is a key part of trial design in MAB problem. On one hand, we should not stick on the SBS with high sample mean since the algorithm may be trapped in a local optimum; on the other hand, continuously trying different SBSs is also not a good idea since it impacts the efficiency of the algorithm. In this paper, we utilize the widely-used UCB policy proposed by [16] to handle this trade-off, since it can achieve logarithmic regret with low computation complexity [8].

According to UCB, we set the index of SBS k in grid g_i as $\bar{r}_i^k(T_i^k) + \theta\sqrt{\frac{2\ln F_i}{T_i^k}}$, where F_i denotes the total number of handovers happened in the grid. The first item acts as the exploitation part, while the second item takes charge of the exploration part with exploration rate θ. For an Event A2 occurring in grid g_i, CC selects the SBS k^* satisfying

$$k^* = \arg\max_{k \in B_i}\left(\bar{r}_i^k(T_i^k) + \theta\sqrt{\frac{2\ln F_i}{T_i^k}}\right). \tag{10}$$

D. Acceleration Technique

Generally, when an Event A2 occurs on a LOS connection which was built in grid g_i to serve UE n by SBS k, a reward r_i^k is obtained and only $\bar{r}_i^k$ will be updated (time- and round- related subscripts are omitted). However, since the UE's trajectory (in terms of the sequence of grids it passes) is known by using gridding scheme, we are able to update some other SBSs' rewards on this trajectory simultaneously, by using the so-called *virtual update*. Specifically, in the previous handover, when CC switched the UE n to the SBS k in grid g_i, CC was also aware of the set of SBSs which were not selected, denoted as $\bar{B}_i^k = B_i \backslash \{k\}$, and pretended to build a virtual LOS link between the UE n and each $k^{'} \in \bar{B}_i^k$. During the UE's post-handover movement, in addition to checking the handover trigger condition for the true LOS link, CC keeps checking that of each virtual LOS link. If the virtual LOS path between the UE n and the SBS k' was blocked, the observed reward $r_i^{k'}$ is calculated and used to update the sample mean $\bar{r}_i^{k'}$, although the corresponding handover event did not truly occur. By this virtual update, any trajectory of UE can be used to update multiple sample means and the efficiency of the algorithm can be improved significantly.

The MAB-based BS selection algorithm with acceleration can be summarized in Algorithm 1.

V. Numerical Experiments

In this section, we compare the performance of two versions of proposed online learning handover policy, without and with acceleration (denoted as MAB and MAB_acc, respectively), with those of two existing handover policies, called RFH and SMART [8], in variant scenarios. In RFH, the BS which provides the maximum transmission rate is chosen as the handover result, and SMART proposes a reinforcement-learning based BS selection algorithm to make handover decisions.

A. Experiment Settings

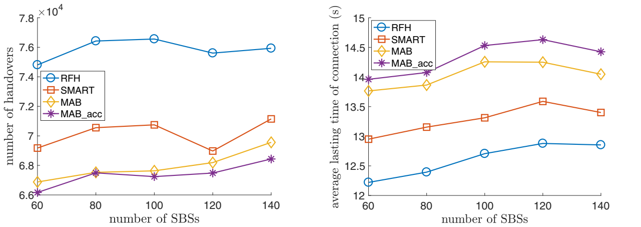
We consider a cellular network N which is built in a 100(m)×100(m) square region and consists of 100 SBSs using mmWave band. The transmit power of SBS is set to be 30 dBm, and the noise power is -77 dBm. Similar to [13], we set the parameters α and β in (1) as 61.4 and 2, respectively. The channel gain G_{max} of main lobe is 18 dB as in [14]. The bandwidth of SBS is set as 500 MHz. We assume that 20 identical obstacles with radius of 1(m) are randomly distributed in the network. According to the gridding scheme and for illustration purpose, the whole network area is partitioned into $20 \times 20 = 400$ identical grids and each grid is a 5(m)×5(m) square area. The signal strengths received at the geometrical center of each grid are calculated and used as the representative RSS vector. The number of UEs entering into the network per time slot has a Poisson distribution with parameter λ. For a new coming UE, its initial position is uniformly distributed at the boarder of the network and its moving orientation is also uniformly distributed. The UE's moving velocity is supposed to be 1(m/s). Any UE's experience is used to update the accumulated reward of SBS until it moves out of the network region. Furthermore, the exploration rate θ is set to be 1.

B. Number of SBSs

In this experiment, we compare the performances of these handover polices with variant numbers of SBSs. We choose five numbers: 60, 80, 100, 120 and 140, and run 20000 iterations for each instance. The arrival rate of UE is set to be 1. The results are shown in Fig. 2. It could be found that, although SMART outperforms RFH, the proposed online learning handover policy performs better than both of them, no matter on the number of handovers or average lasting time of connection. Compared with RFH, the online learning handover policy can improve the number of handovers and the average lasting time of connection by 8% – 11% and 9% – 13%, respectively. Due to acceleration, the online learning handover policy with virtual update performs the best over all polices.

Algorithm 1: MAB-based BS Selection Algorithm

```
Input: Cellular network $N$ with gridding, consisting of
       mmWave small cells (system parameters, a set of
       SBSs, a set of UEs, a set of obstacles)
Output: SBS selection result $k^*$
Initialization: For all $g_i \in N$ and $k \in B_i$, $T_i^k \leftarrow 0$,
  $\bar{r}_i^k(0) \leftarrow 0$, $F_i \leftarrow 0$;
while Event A2 handover trigger condition is met for a
  UE $n$ do
    Get the current time slot $\tau$ and identify the grid $g_j$
      where the UE $n$ currently resides;
    Load the serving SBS $k$, the grid $g_i$ where and the
      time $\tau_0$ when the current connection was built;
    Calculate the reward $r_{i,\tau_0}^k = \tau - \tau_0$ for the SBS $k$;
    $\bar{r}_i^k(T_i^k + 1) \leftarrow \frac{T_i^k \times \bar{r}_i^k(T_i^k) + r_{i,\tau_0}^k}{T_i^k + 1}$;
    $T_i^k \leftarrow T_i^k + 1$;
    $F_i \leftarrow F_i + 1$;
    Update $\bar{r}_i^{k'}(T_i^{k'} + 1)$, $T_i^{k'}$ and $F_i$, for $k' \in \bar{B}_i^k$;
    $k^* = \arg\max_{k \in B_j} \left( \bar{r}_j^k(T_j^k) + \theta \sqrt{\frac{2 \ln F_j}{T_j^k}} \right)$;
    Switch the UE $n$ to the SBS $k^*$;
end
```

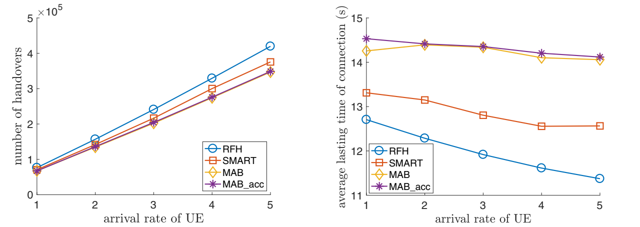


(a) Performance on the number of handovers (b) Performance on the average lasting time of connection

Fig. 2. Comparison of performances with different numbers of SBSs

C. UE's Arrival Rate

In this experiment, we compare the performances of these handover polices with different arrival rates of UE. The arrival rate of UE reflects the crowdedness of the network. In order to simulate different practical scenarios with different degrees of crowdedness, we choose five values for λ: 1, 2, 3, 4 and 5, and run 20000 iterations for each instance. The number of SBSs is fixed to be 100. The results are displayed in Fig. 3. As shown, with all the chosen values of λ, the proposed online learning handover policy always brings the fewest number of handovers and the longest average lasting time of connection. Moreover, with the growth of arrival rate, the online learning handover policy's advantage over the other two policies, saying RFH and SMART, becomes more remarkable. This means that the online learning handover policy is more prominent even in crowded situation.



(a) Performance on the number of handovers (b) Performance on the average lasting time of connection

Fig. 3. Comparison of performances with different arrival rates

D. Regulation of UE's Movement

In the above experiments, the UE's initial position and orientation are totally random without any constraint. However, in many real-world scenario, specially in urban area, the UE's movement is highly restricted. Specifically, since the UE can only move along with the existing sidewalks, the UE's position is actually limited within the area of the sidewalk but not the whole network, and the UE's orientation is directed by the direction of sidewalk. In this experiment, we add four sidewalks into the cellular network in order to make our simulation closer to reality, as the areas with shadow shown in Fig. 4. New UE is generated at a border of a sidewalk and moves to the opposite boarder. Besides, we introduce a specific parameter $\gamma \in [0, 1]$ to describe the homogeneity of the UEs' movements within the sidewalks. In particular, if $\gamma = 0$ or 1, all UEs in the same sidewalk move with the same orientation; if $\gamma = 0.5$, half of the UEs move oppositely. The number of SBSs and the arrival rate of UE are set to be 100 and 1, respectively. The other parameters keep the same with the above experiments. We set different values to γ and run 20000 iterations for each instance. The results are shown in Fig. 5. It could be found that, RFH performs the worst and its performance almost does not change with the growth of γ. SMART is better than RFH, and its performance slightly fluctuates with different values of γ. The online learning handover policy still performs the best and its advantage over the former two polices is more significant than those in the scenarios with randomly moving UE which are shown in the above two experiments, although its performance decreases a little when γ approaches to 0.5. This means that, the online learning handover policy is more suitable for the scenario where UE has regular movement.

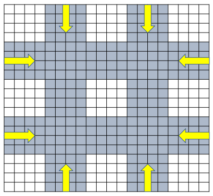
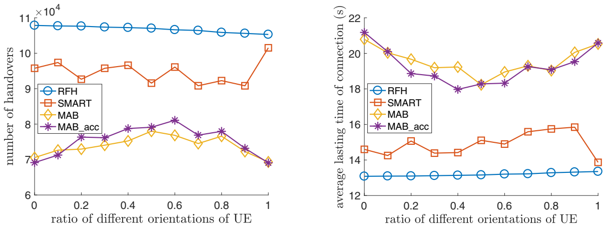


Fig. 4. Scenario with sidewalks



(a) Performance on the number of handovers

(b) Performance on the average lasting time of connection

Fig. 5. Comparison of performances on different orientations of UE

VI. Conclusions

In order to reduce unnecessary handovers in ultra-dense mmWave cellular network, we propose that the empirical knowledge extracted from historical handovers is of benefit for wisely choosing the optimal BS as handover decision, which can bring long user-BS connection time after handover. A novel handover mechanism which exploits the empirical distribution of UEs' post-handover trajectory and LOS blockage to guide future handover without any exact location information of UE is presented. In particular, we design a gridding scheme to utilize the RSS information as an index to retrieve the accumulated rewards of candidate BSs. These accumulated rewards are treated as empirical knowledge which is learned online through a multi-armed bandit framework. An effective BS selection algorithm with acceleration technique is also proposed. The results of numerical experiments show that the proposed online learning handover policy outperforms existing counterparts in variant real-world scenarios significantly.

Acknowledgment

The work of T. Shu is supported in part by NSF under grants CNS-1837034, CNS-1745254, CNS-1659965, and CNS-1460897. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of NSF.

References

[1] A. Talukdar, M. Cudak, and A. Ghosh, "Handoff rates for millimeterwave 5G systems," in Proc. IEEE 79th Veh. Technol. Conf., 2014, pp. 1–5.

[2] H. Shokri-Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, "Millimeter wave cellular networks: a MAC layer perspective," IEEE Transactions on Communications, vol. 63 , no. 10, pp. 3437–3458, 2015.

[3] M. Giordaniy, M. Mezzavilla, S. Rangan, and M. Zorzi, "Multi-connectivity in 5G mmWave cellular networks," Mediterranean Ad Hoc Networking Workshop, 2016.

[4] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved handover through dual connectivity in 5G mmWave mobile networks," IEEE Journal on Selected Areas in Communications, vol. 35, no. 9, pp. 2069–2084, 2017.

[5] O. Semiari, W. Saad, M. Bennis, and B. Maham, "Mobility management for heterogeneous networks: leveraging millimeter wave for seamless handover," IEEE Global Communications Conference, 2017.

[6] M. Mezzavilla, S. Goyal, S. Panwar, S. Rangan, and M. Zorzi, "An MDP model for optimal handover decisions in mmWave cellular networks," European Conf. on Networks and Communications, 2016, pp. 100–105.

[7] S. Zang, W. Bao, P. L. Yeoh, H. Chen, Z. Lin, et al., "Mobility handover optimization in millimeter wave heterogeneous networks," International Symposium on Communications and Information Technologies, 2017.

[8] Y. Sun, G. Feng, S. Qin, Y.-C. Liang and T.-S. P. Yum, "The SMART handoff policy for millimeter wave heterogeneous cellular networks," IEEE Trans. on Mobile Comput., vol. 17, no. 6, pp. 1456–1468, 2018.

[9] A. Alkhateeb and I. Beltagy, "Machine learning for reliable mmWave systems: blockage prediction and proactive handoff," IEEE Global Conference on Signal and Information Processing, 2018.

[10] H. Okamoto, T. Nishio, M. Morikura, K. Yamamoto, D Murayama, et al., "Machine-learning-based throughput estimation using images for mmWave communications," IEEE Vehicular Technology Conf., 2017.

[11] B. V. Quang, R. V. Prasad, and I. Niemegeers, "A sruvey on hand-offs lessons for 60 GHz based," IEEE Commun. Surveys Tutorials, vol. 14, no. 1, pp. 64–86, 2012.

[12] 3GPP TS 36.331, "E-UTRA radio resource control (RRC); protocol specification (Release 9)," 2016.

[13] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, et al., "Millimeter wave channel modeling and cellular capacity evaluation," IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 1164–1179, 2014.

[14] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," IEEE J. Sel. Areas Commun., vol. 33, no. 10, pp. 2196–2211, 2015.

[15] S. Maghsudi and E. Hossain, "Multi-armed bandits with application to 5G small cells," IEEE Wireless Communications, vol. 23, no. 3, pp. 64–73, 2016.

[16] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," Mach. Learning, vol. 47, no. 2/3, pp. 235–256, 2002.