# Spatial and Temporal Contextual Multi-Armed Bandit Handovers in Ultra-Dense mmWave Cellular Networks

Li Sun, Jing Hou, and Tao Shu

**Abstract**—Although millimeter wave (mmWave) is a promising technology in 5G communication, its severe path attenuation and susceptibility 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. Since conventional handover policies purely rely on the measurement of signal strength, 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. In this paper, we propose two novel handover mechanisms to reduce unnecessary handovers 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. Without prior knowledge of user's mobility and environment, the proposed handover mechanisms exploit the empirical distribution of user's post-handover trajectory and LOS blockage, learned online through a multi-armed bandit (MAB) framework. Depending on the contexts extracted from RSS information, two different MAB problems for handover are formulated, which focus on spatial and space-time contexts, respectively, The results of numerical simulations demonstrate that the proposed contextual handover mechanisms significantly outperform existing counterparts on reducing handovers in all simulated scenarios.

**Index Terms**—millimeter wave, ultra-dense cellular network, handover management, spatial and space-time context, multi-armed bandit.

## 1 Introduction

As one of the fundamental technologies in the upcoming 5G cellular networks, millimeter wave (mmWave) can provide abundant bandwidth for wireless service through the line-of-sight (LOS) path because of its 10-to-100s GHz level frequency. 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 is susceptible to blockage 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]. Moreover, in an ultra-dense network, when handover is needed, there are typically multiple candidate BSs that could be chosen to handover to. Therefore, efficient mobility and handover management is an inherent challenge that needs to be addressed in ultra-dense mmWave cellular networks.

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 [2]. 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.

The study on reducing unnecessary handover in mmWave cellular networks has been just started. The authors in [3] introduced a handover strategy based on report tables that were built by BSs to track the stability of their surrounding channels. By analyzing the report tables, the decision maker could make a good handover decision by choosing the BS with high stability of channels to avoid a possible handover again in the very near future. The authors in [4] developed a Recursive Least Square (RLS) based algorithm to predict the received signal strength (RSS) of BS and chooses the BSs with the largest predicted RSSs. Some works attempted to reduce the handover frequency by employing statistical predictive models such as finite

*A preliminary version of this work has been presented in IEEE Globecom 2019, Waikoloa, HI, in Dec. 2019.*

*Li Sun, Jing Hou and Tao Shu are with the Department of Computer Science and Software Engineering, Auburn University, Auburn, AL 36849, USA, e-mail: {lzs0070, jzh0141, tshu}@auburn.edu.*

state Markov chain [5] and Markov decision process (MDP) [6] to predict the possibility of an outage in the next time slot based on the current channel state. Some others proposed to solve this problem by utilizing machine-learning based frameworks [7]–[10]. Furthermore, [11] developed a geometry-based blockage prediction method to eliminate unnecessary handovers caused by short-term LOS blockage.

Although the above works make outstanding contributions in this field and provide enlightenment to our study, there is still space to improve. First, many existing handover strategies, especially those based on statistical predictive models (e.g., MDP [6]), require the pre-knowledge or assumption on the distribution of the channel state, so there is no guarantee that they can be readily applied in a ultra-dense cellular network [7]. Second, many strategies have specific requirements, e.g., antenna array equipped at UE and exhaustive direction search [3], assumption of known mobility of UE [8] and time-consuming offline training [10]. Third, few of these strategies explicitly consider the contextual relationship between LOS link, user's movement, and obstacles. Clearly, the unobstructed time for a LOS link is essentially determined by user's movement trajectory and the distribution of LOS blockage after the handover. Therefore, a handover policy could have been improved by exploiting user's post-handover mobility trajectory and LOS blockage, but the issue is that the realization of these information 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 her trajectory before the handover. However, this solution requires exact location information of the user (i.e., geo-coordinates of UE's location), which is not always available/practical in reality.

In this paper, we propose two handover mechanisms that addresses these challenges by exploiting the *empirical distribution* of user's post-handover trajectory and LOS blockage. This empirical knowledge orientated handover mechanisms are based on the following logic: if a handover policy is optimal for a handover, then it is quite likely to be also optimal for other handovers of the same features. Here, the *features*, which describe the context or environment of user's communication, are defined by some representative characteristics related to user's post-handover trajectory and LOS blockage. The feature acts as a label to group similar handovers, to which a common handover decision optimal to this group will be applied. Taking into account several fundamental attributes of handovers, including the user's mobility, the environment, and the consequent unobstructed time for LOS path, we develop a novel *partitioning scheme* to extract key features in the *space* and *time* domains based on the UE's RSS information. With the assistance of these features, the proposed handover mechanisms depending on the availability of the features can significantly increase the lifetime of the LOS link after each handover without requiring any exact location information of the users.

In our mechanisms, the empirical knowledge is learned *online* through a multi-armed bandit (MAB) framework, with the intention to maximize the expectation of the unobstructed time for user-BS connection after each handover. In particular, the centralized controller of the cellular system maintains an individual MAB process for each block. A block, which indicates a specific Euclidian area in the network, is used as the spatial feature of handover (definition will be given shortly). All UEs within the same block will see the same set of BSs. These available BSs are treated as the arms of the MAB process associated with the block, and choosing a BS as the handover destination is viewed as a play. When a handover is triggered, the system firstly identifies the block in which the UE resides, so as to identify the particular MAB process associated with that block. It then chooses a BS as the UE's handover destination among all available BSs of that MAB according to their accumulated rewards. Handing over to this BS, the UE will receive an instantaneous reward (definition will be clear shortly) from its communication with the BS. This reward will be reported to the system by the UE at the moment of its next handover. This reward is used to update the accumulated reward of the serving BS for the corresponding MAB process, which is used to guide future handovers of the same spatial feature. We propose two BS-selection algorithms to 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 algorithms. In contrast to the aforementioned trajectory prediction method, an advantage of our mechanisms is that they do not require user's exact location information. Instead, user's coarse-grained mobility information, which differentiates handovers with the areas where the handovers occur and the user's general moving directions, is used as *context* in our algorithm to collect rewards. In practice, this coarse-grained information could be represented as the collection of RSSs from surrounding BSs, and hence is considered practical according to 3GPP [12]. Note that the purpose why we introduce to use a collection of RSSs is to differentiate the UEs with different features, but not really figure out their exact geographic locations or moving directions.

In the literature, our work is most related to the SMART scheme [7] and a similar one [8], which also use 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 various features of handover to better characterize the accumulated knowledge, while SMART is completely independent from user's individual characteristics. Our performance evaluation simulates SMART as a counterpart scheme and shows that the proposed mechanisms outperform SMART significantly. In addition, [13] also proposed a MAB-based algorithm to optimize wireless handover problem. Our work differs from it in the following two aspects. First, the handover problem discussed in [13] is optimized by fine tuning the threshold defined in event A2, i.e., a handover will be triggered whenever the received signal strength goes below a pre-defined threshold [12]. This traditional setting is not suitable for mmWave cellular network due to its special propagation property. Second, the cellular network considered in [13] consists of non-overlapped cells. When an event A2 occurs, the UE has only one target BS to which it can switch and there is no need to choose among multiple candidate BSs. This setting does not consider the property of ultra-dense network. In particular, with overlapped cells, how to choose proper BS among multiple candidates is an important decision. Therefore, the method proposed in [13] cannot solve the problem considered in our work. Moreover,

although [14], [15] considered similar context in handover management, they had their own limitations. Specifically, the handover method proposed in [14] focuses on elaborately tuning handover parameters and is not suitable for handover in mmWave band, just like that in [13]. Besides, this method requires auxiliary devices to collect user's precise speed information. Although [15] proposed a context-aware handover policy without using any positioning systems, the aim of their work is to avoid exhaustively beam-searching to reduce handover delay, which is different with ours. Finally, this paper is an extension of our preliminary work in [16], which only learns and exploits the spatial information of past handovers. In contrast, this paper comprehensively considers both the spatial and *temporal* features of handovers to provide better solutions to the handover problem in ultra-dense mmWave cellular networks.

The rest of this paper is organized as follows. In Section 2, an overview of the related works are proposed. Section 3 introduces the models of channel propagation, the blockage and the mobility. In Section 4, the online learning framework of the two contextual handover mechanisms is described. Section 5 and Section 6 propose two MAB-based BS-selection algorithms for these two handover mechanisms, respectively. In Section 7, we give the complexity analysis of the proposed mechanisms. In Section 8, the results of a series of simulations in various scenarios and discussions are provided. Finally, we conclude our work in Section 9.

## 2 Related Works

Our work is related to two fields: (1) handover strategies for mmWave cellular networks, and (2) machine learning based methods for handover in wireless communication.

### 2.1 Handover Strategies for mmWave Networks

Research on handover management in mmWave cellular networks is in its infancy and the results are preliminary. With the aim to compensate the large propagation path loss and high susceptibility to blockage, existing methods include multiple (parallel) connectivity and single sequential connectivity. The former maintains simultaneous beamforming from multiple base stations to a user, so that the user is still under cover if its LOS to one base station is lost [3], [4], [17]–[19]. The latter beamforms to the user from a single base station at a time, but will handover to a next base station when the current connection is lost [20]–[24]. None of these methods provide a satisfactory solution to the problem. The multiple connectivity method has low efficiency in the beam utilization, because not every beam that has been allocated to the user is always needed to maintain the connection at all time. Meanwhile, the method also suffers from multi-fold user capacity loss, as each user now requires $N$ beams, which could have been used to serve $N$ users. On the other hand, the single sequential connectivity method suffers from long handover delay, since the initial access of the new beam requires expensive signaling and long training time [25]. Some works improve the handover delay by employing statistical predictive models such as finite-state Markov chain [14], [26] and Markov decision process (MDP) [6], [27] to predict the possibility of an outage in the next time slot based on the current channel state. Moreover, [15] introduced a linear-regression based direction of pass detection algorithm to reduce handover delay. In addition, content caching technique has been utilized to lower handover failure rate and smooth handover [28]–[30].

### 2.2 Machine Learning Based Methods for Handover

Machine learning provides another promising tool to improve handover decision. In particular, the authors in [31] introduced a partially blind handover scheme that uses embedded XGBoost classifier to predict the success rate of handover. In [32], the authors employed deep learning (DL) framework to predict upcoming failure events and implement proactive handover based on historical beamforming vectors. However, they did not describe how it works in multi-user scenario. In [33], the authors built a convolutional neural network to predict the signal power that will be received in a short time. But their solution relies on costly camera device, which is not scalable in practice.

Moreover, the authors in [7] introduced a reinforcement-learning (RL) based handover policy to reduce the number of handovers in HetNet. In [34], the authors utilized RL to predict user's mobility and applied proactive handover to improve the throughput. However, their method needs user's velocity and location information obtained via a dedicated tracking device. In [8], the authors considered a communication system consisted of users and unmanned aerial vehicles (UAVs), and proposed a user association algorithm based on RL to reduce redundant handovers.

As an integration of DL and RL, deep reinforcement learning (DRL) is also utilized to reduce handover frequency in wireless communication. In [9], the authors proposed an asynchronous multi-user DRL scheme with a deep neural network (DNN) as handover controller to reduce handover frequency. This scheme requires user's geographical information and uses this information as the feature to partition UEs by $K$-means clustering algorithm. Similarly, the authors in [10] proposed a handover scheme based on deep $Q$-network. Different from [9], [10] utilized the historical received uplink SINR on APs to characterize the UE's state and leveraged the convolutional neural network and the recurrent neural network to extract UE's features.

### 2.3 Discussion

Upon the plentiful research results in the related fields, the limitations of the existing works and the specific contributions of ours can be summarized as follows:

- Few of the existing research explicitly considers the impact of distributions of user's mobility and LOS blockage on handover frequency in mmWave cellular networks. Most research in the literature uses throughput [7], [14], [17], [19], [34], or delay [6], [15], [26], [27], or failure rate [18], [28]–[32] as the evaluation criteria for handover policy. Rather than these metrics, we focus on the unobstructed time for a LOS link which more directly reflects the quality of a handover decision due to directivity of mmWave communication. The estimation of the unobstructed

LOS time requires certain knowledge of user's post-handover trajectory and LOS blockage, whose acquisition has not been studied in the literature.
- Most of the existing solutions have specific requirements, for example, the prior knowledge on the distribution of channel state [6] and user's mobility [9], [14], [27], [31], or auxiliary devices to obtain user's movement information [13], [33], [34]. However, such requirements cannot always be satisfied in practice. Hence, we propose two novel handover mechanisms which leverage the available RSS information to extract user's spatial and temporal features to guide the handover decision without any pre-knowledge on user's exact mobility information.
- We propose an online learning framework based on MAB process with low computational complexity. This online learning has a simple structure which requirs no offline training phase or hyper-parameter tuning, hence is easy to implement.

In summary, we propose two novel online-learning-based contextual handover mechanisms, which can learn the empirical distribution of user's post-handover trajectory and LOS blockage, and use the learning outcome to reduce unnecessary handovers in ultra-dense mmWave cellular networks. Depending on the availability of information, two different MAB formulations are proposed for the learning, one focused on features of handover events in the space domain, and the other on features in both the space and the time domains. Two effective BS-selection algorithms are developed for these two mechanisms, respectively. Moreover, in order to address the issue caused by high-dimension feature of handover in a complex scenario, a novel acceleration technique is presented to increase the efficiency of the algorithms. Note that, none of these proposed mechanisms requires the knowledge of user's exact (or fine-grained) mobility information.

## 3 System Model

Consider a cellular network N consisting of a set of mmWave small cell base stations (SBSs), denoted as $\mathbf{S}$. 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 its conventional sub-6 GHz band, it is used for transmission of control 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 the 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.

### 3.1 Propagation Model

In this paper, we assume that the channel of a mmWave SBS is described by 3GPP Standard probabilistic LOS model. According to [7], [35], the statistic path loss model is

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

where $d$ is the distance between the transmitter and the receiver in meters, $\alpha$ and $\beta$ are the least square fittings of floating intercept and slope respectively over the measured distances, and $\xi$ represents a lognormal shadowing with variance $\sigma^2$. Since inter-user interference can be ignored in mmWave band, we only model the signal-to-noise ratio (SNR) of the signal received by the UE $n$ from the SBS $k \in \mathbf{S}$ as [7]

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

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 antenna gain $G$ is actually a function of the angle of departure $\omega$ from the SBS to the UE. According to [36], this function can be represented by

$$G(\omega) = \begin{cases} G_{\max}, & \text{if } |\omega| \leq \omega_{\mathrm{s}} \\ G_{\min}, & \text{otherwise,} \end{cases} \quad (3)$$

where $G_{\max}$ is the main lobe gain, $G_{\min}$ is the side lobe gain, and $\omega_{\mathrm{s}}$ is the main lobe width of the SBS. We assume that perfect beam tracking technique can be used to maintain mmWave link [7]. Therefore, the UE could always be in the main lobe and have main lobe gain as long as its LOS path to the SBS is not blocked.

We assume that a SBS is able to generate at most $U_{\max}$ beams at the same time (i.e., it can transmit to at most $U_{\max}$ UEs simultaneously), and all served UEs equally share its bandwidth. The downlink transmission rate of a UE $n$ that a SBS $k$ is transmitting to can be calculated as follows:

$$h_n^k = \frac{B_{\mathrm{w}}}{U_k} \log_2(1 + SNR_n^k), \quad (4)$$

where $B_{\mathrm{w}}$ is the bandwidth of SBS $k$ and $U_k$ is the number of UEs simultaneously served by SBS $k$.

### 3.2 Blockage and Mobility Model

Since mmWave signal is highly vulnerable to obstacles, we assume that the transmission rate of a LOS link will drop to zero immediately when the link is blocked. In our simulations, a user is modeled with random moving speed and direction, while an obstacle is modeled as a circle with fixed radius and location. Given a LOS link and a set of randomly distributed obstacles, the link is blocked whenever there is an obstacle to which the distance from the link is less than its radius. This modeling could be used to represent any fixed obstacle, such as tree-top, advertising board and building. Note that our proposed mechanisms and their analysis do not make any assumption on the blockage and mobility models. In other words, they are general enough to work under any blockage and mobility models that may appear in realistic applications. The main reason why we select this model is because it is easy to simulate but still general enough. Even though we assume a unified radius for all obstacles in our simulations, this assumption does not undermine the generality of the obstacle model in the sense that the blockage time caused by an arbitrary obstacle to an

arbitrary user in our model is still a random variable. This is because this blockage time depends on not only the size of the obstacle, but also the distance between the obstacle and the user, and the user's moving direction and speed, which are randomly distributed in the model. Therefore, this model has already been able to capture the fundamental effects of heterogeneous blockage time that could have been caused by a more complicated obstacle model. Moreover, in this paper we mainly consider the static-obstacle and mobile-user scenario. The more challenging mobile-obstacle scenario is out of the scope of this paper and will be considered in our future work.

Moreover, we do not consider any communication through non-LOS (NLOS) in this work. Although recently there have been commercial tests, e.g., those from Qualcomm [37], Samsung [38], and NI [39], that show the feasibility of using NLOS for communication when there is no LOS, simply switching the beam to a NLOS component of the mmWave channel when the LOS is blocked [40]–[43] may not always be a good solution to our handover problem. In particular, recent field measurements in New York City by NYU have shown that for all the frequencies of practical interest in the mmWave cellular band (28, 38, and 73 GHz), the strength of a NLOS is in general at least 20 to 30 dB weaker than that of the LOS at typical outdoor communication ranges. Due to the huge difference in the path loss between LOS and NLOS, switching beams to NLOS may either lead to a transmission rate that is orders of magnitude lower than the LOS rate if transmission power is not increased, or cause a huge spike in power consumption if one wishes to retain a comparable transmission rate. Clearly, rather than simply switching to the NLOS, a more sophisticated cross-layer mechanism that can mobilize the rich networking resources embedded in the dense deployment of mmWave cells, e.g., by handover when possible to another base station or a relay that has a new LOS with the user, may avoid the above weaknesses and thus constitutes a more desirable solution. Hence, we do not consider the effect of reflected signal by obstacles in this work.

## 4 Online Learning of Contextual Handover Mechanisms

Within the six handover events defined by 3GPP standard, we focus on the BS selection for Event A2 (i.e., a handover will be triggered whenever the received signal strength goes below a pre-defined threshold [12]), since handover triggered by A2 is common but challenging in mmWave band. At a high level, the framework of the proposed MAB-based online learning of contextual handover mechanisms is illustrated in Fig. 1, and elaborated in the following.

### 4.1 Spatial Contextual Handover Mechanism (SCH)

#### 4.1.1 Signal Space Partitioning Scheme

During a handover, instead of picking the BS that has the highest instantaneous RSS, we prefer a BS that has the longest connection time for its LOS link subject to a minimum RSS requirement. The unobstructed time of LOS link is determined by user's 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 another word, the geographic area where a handover occurs should be considered as a *spatial feature* of the handover when making BS selection. Unfortunately, this spatial feature of handover cannot be directly obtained without precise location information, i.e. geo-coordinate. In order to observe this spatial feature without any auxiliary locating device, in this subsection, we introduce a *signal space partitioning scheme*, which leverages the UE's RSS information as a label to characterize each handover. This setting allows handovers to be differentiated by the areas where they occur.

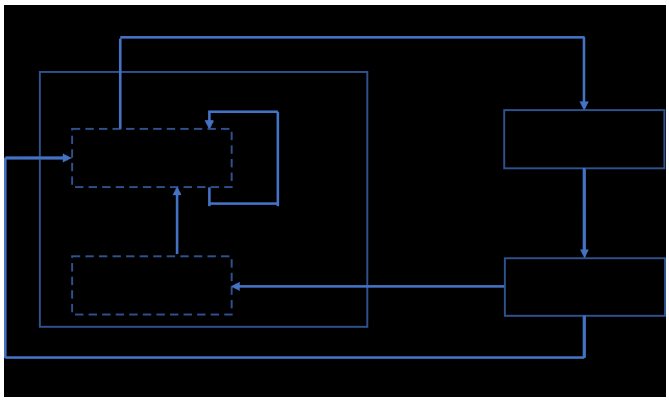


Fig. 1. Framework of online learning of contextual handover mechanisms

The idea of this scheme comes from the observation that the collection of signals received from surrounding SBSs could be leveraged as a reference of UE's location. Specifically, in an ultra-dense 5G network, it is common that multiple mmWave small cells overlap. Therefore, a UE $n$ at any location is likely to receive from multiple surrounding mmWave SBSs, which form the available BS set $\mathbf{S}_n$. All SBSs $k \in \mathbf{S}_n$, associated with the $SNR$s at the UE received from them, constitute a *signal vector* for UE $n$, denoted by $\boldsymbol{v}_n$. Each entry $v_k \in \boldsymbol{v}_n$ is a quantized version of the $SNR_k$ received from SBS $k$ according to the following quantizating criterion: choose $J$ quantizing thresholds $\{e_0, \ldots, e_{J-1}\}$, where $J$ is a parameter and $e_{j_1} < e_{j_2}, 0 \leq j_1 < j_2 \leq J-1$, then define the quantized SNR as

$$v_k = \begin{cases} J, & \text{if } SNR_k \geq e_{J-1}, \\ j, & \text{if } e_{j-1} \leq SNR_k < e_j, 1 \leq j < J-1, \\ 0, & \text{if } SNR_k < e_0. \end{cases} \quad (5)$$

In this way, any instance of signal vector $\boldsymbol{v}_n$ corresponds to a certain geographic area where UE $n$ instantaneously locates. We use an example shown in Fig. 2 to illustrate the idea.

In Fig. 2, we consider three SBSs, A, B and C, whose small cells overlap. Note that the small cell of a SBS, which is represented by a dotted circle, indicates the region within which a UE is able to receive from the SBS, i.e., the RSS is above the minimum required threshold. So here we are considering a binary quantization case where a single quantizating threshold $e_0$ exists. There are 9 UEs distributed in the network. If a UE is in the small cell of a SBS and there is no blockage on the LOS between the UE and that SBS, the UE's quantized SNR corresponding to the SBS is 1, otherwise 0. In this setting, all UEs' signal vectors are listed in Table 1. It is easy to see that UEs at different locations receive different signal vectors, hence have different spatial

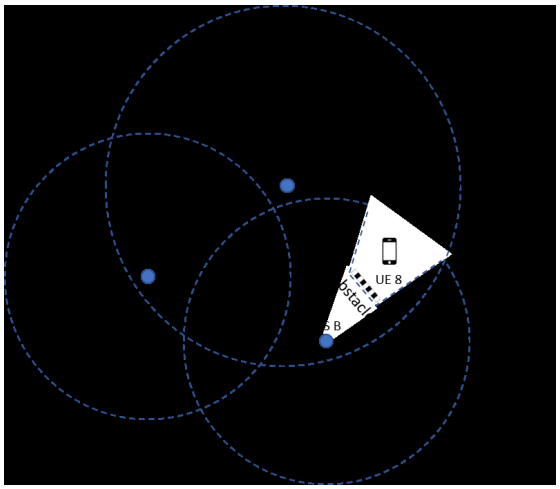



Fig. 2. Illustration of signal space partition

TABLE 1
Signal vectors

| UE | Quantized SNR | | | Signal Vector |
|---|---|---|---|---|
| | A | B | C | |
| 1 | 1 | 0 | 0 | 1 0 0 |
| 2 | 0 | 1 | 0 | 0 1 0 |
| 3 | 0 | 0 | 1 | 0 0 1 |
| 4 | 1 | 1 | 0 | 1 1 0 |
| 5 | 1 | 0 | 1 | 1 0 1 |
| 6 | 0 | 1 | 1 | 0 1 1 |
| 7 | 1 | 1 | 1 | 1 1 1 |
| 8 | 1 | 0 | 0 | 1 0 0 |
| 9 | 0 | 0 | 0 | 0 0 0 |

features. In particular, note that UE 8 does not have the same signal vector as UE 4, even though both of them reside in the overlap between the small cells of SBS A and SBS B. This disparity in received signal vector arises from the blockage of the LOS between SBS B and UE 8 caused by the obstacle. Instead, UE 8 has the same signal vector as UE 1, also due to the blockage of the obstacle. As a result, UE 1 and UE 8 are considered to have the same spatial feature in our model.

According to our signal space partitioning scheme, whenever a handover is triggered, the CC collects the UE's instantaneous RSS and identify its signal vector which indicates the area where the handover event occurs. Indeed, the proposed partitioning takes place in the signal space. Because the signal path loss is related to signal propagation distance, the partition in the signal space naturally leads to a partition in the Euclidian space, referred as *blocks*. In this way, each signal vector corresponds to a unique block. Rather than directly identifying the specific location of a UE, the main goal of the proposed partitioning scheme is to identify UEs with the same spatial feature (i.e., residing in the same block) by assigning them the same block Id (i.e., the Id of the block that the UE is currently residing in). Different block Ids will be assigned to the UE as it moves and receives different signal vectors.

The information about the blocks, such as the amount and the size, referred as the granularity of partition, is determined by the chosen quantization thresholds as well as the distribution of SBS and obstacles. This partitioning scheme does not rely on any pre-knowledge or assumption on these blocks. Instead, the knowledge on the blocks is grown incrementally. In particular, the CC keeps a block set storing the Ids of the blocks that have been identified. Initially, the block set is empty. Once the CC receives a new signal vector from a UE that has never been observed before, this means a new block is identified, and the CC assigns a unique Id to the new block and add it into the block set. When more blocks are identified, the CC will have more complete knowledge on the blocks.

#### 4.1.2 Spatial Contextual BS-Selection based on Empirical Knowledge of Post-handover Trajectory

In SCH mechanism, given $M$ blocks (each corresponding to a signal vector), the CC maintains $M$ independent MAB processes, each serving a block by selecting SBS for handover events happening in that block. In particular, suppose a UE in block $g_i \in \mathbf{G}$, where $\mathbf{G}$ denotes the set of blocks, is able to receive from $n_i$ SBSs, denoted by set $\mathbf{S}_i$. Then the MAB process for block $g_i$ has $n_i$ arms, each representing a distinct SBS in $\mathbf{S}_i$. 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 block $g_i$, so it reflects the mean reward a future UE is expected to receive if it switches to SBS $k$ after a handover in block $g_i$.

The spatial feature of handover provides the CC an access to coarsely locate each handover in the network. Different with the handover mechanism in [7] which treats all handovers with no difference and applies a unique RL process, the SCH maintains multiple MAB processes and utilizes the spatial feature of each handover as a context to indicate which MAB process should be enabled. With the MAB process corresponding to a specific block, an incoming UE that handovers in that block will be switched 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 the 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 block $g_i$ is based on all historical realizations of UE's post-handover trajectories for handovers in $g_i$, and hence it is an expectation over the empirical distribution of UE's post-handover trajectory.

### 4.2 Space-Time Contextual Handover Mechanism (STCH)

#### 4.2.1 Temporal Feature Extraction

The UE's mobility is reflected by not only its instantaneous location, but also its moving direction. A single block Id identified by the signal space partitioning scheme can only label out the coarse-grained location of UE, but cannot indicate the moving direction. Hence, we propose to use a sequence of block Ids, i.e., a block concatenation, as a label to identify UE's moving direction. In particular, the CC maintains a block concatenation for each UE, which records the blocks that the UE has passed in chronological order. It

reflects the change of the UE's location over time and hence can be used as a label representing the UE's coarse-grained moving direction (or trajectory) in the past. In a short time horizon, a UE's moving direction in the near future should be closely related to its moving direction in the past. This correlation between the near future and the past is the basis for the STCH to make a better handover decision.

Take the handover events shown in Fig. 3 as an example, where we suppose UEs 1, 2 and 3, each with its own moving direction indicated by the corresponding arrow, have to hand over in block a. We also assume that all three UEs are able to receive adequate signals from SBSs A, B and C at this moment. Under the SCH mechanism, these three UEs receive the same signal vector and will be treated homogeneously, and therefore will all be handed over to the same SBS, e.g., say SBS C. Clearly, this is not the best handover decision when UE's moving direction is concerned. In particular, as UE 2 will be subsequently getting away from SBS C and getting closer to SBS B, handover to SBS B should be a better decision for UE 2. This could save UE 2 from an (unnecessary) handover from SBS C to SBS B in a later time if it chooses to handover to SBS C at this moment. Similarly, handover to SBS A is a better choice for UE 3, as it could save the UE from an (unnecessary) handover from SBS C to SBS A in the near future if the UE hands over to SBS C in the first place. Using the concatenation of historical block Ids as a label allows us to better classify those UEs with the same spatial feature but are moving along different directions according to their pre-handover trajectories, and hence offers an opportunity to better tailor the handover decisions for them. For example, with the block concatenation of each UE, we are able to know that UE 1, 2 and 3 come from block d, b and c, respectively. By considering their past moving directions, and also based on the learned reward statistics for each handover option related to each past moving direction, the CC could hand over these UEs to SBSs C, B and A, respectively, and hence a better handover decision for the UEs. Note that the introduction of block concatenation is not to really predict UE's moving direction, but to differentiate UEs with different moving directions according to their pre-handover trajectories. Although this concatenation-based moving direction is coarse-grained, it still provides us valuable information to further differentiate UEs with the same spatial feature. Obviously, with finer partition granularity of the signal space and longer block concatenation, the concatenation-based moving direction is more differentiable. As will be clear shortly, even with a very simple one-step-look-back label construct, i.e., each concatenation only contains the current block Id and the most recent one the UE just traversed, the handover performance can be significantly improved.

In general, the concatenation of historical block Ids can be implemented as a stack, where the bottom stores the Id of the first block which the UE has traversed, while the top stores that of the block where the UE currently resides. Specifically, let $\mathbf{g}_n$ denote the block concatenation of UE $n$. Whenever the UE enters a different block, for example, moving from block $g_j$ to $g_i$, the CC pushes $g_j$ into $\mathbf{g}_n$. When UE $n$ needs to handover, the $\mathbf{g}_n$ is used to indicate its moving direction. The block concatenation $\mathbf{g}_n$ extends the observation of handover from the space domain to the time domain, and is referred to as the *temporal feature* of the handover. Because maintaining the entire block-traversing history for every UE is expensive, we will only focus on its simplest one-step-look-back special case, which has low cost and is easy to implement. In particular, for the one-step-look-back construct, let $\tilde{\mathbf{g}}_n = (g_n^-, g_n)$, where $g_n$ denotes the block where UE $n$ currently resides, while $g_n^-$ denotes the immediate preceding block to the current one.

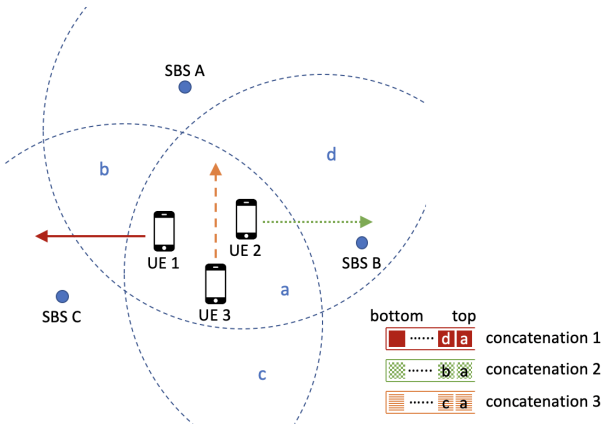



Fig. 3. Differentiation of UEs' moving directions

There are mainly two reasons for us to apply this one-step-look-back construct: (1) Avoid the "curse of dimensionalit". Since every distinct permutation of block Ids could be treated as an instance of block concatenation which is used as an individual space-time context of a handover, if the size of block concatenation is large, there will be a huge number of contexts for a single block. For example, suppose each block has average $p$ preceding blocks and the size of block concatenation is $L$, there are totally $\mathbf{O}(p^{L-1})$ distinct space-time contexts for each handover, which are not manipulable. However, if the one-step-look-back construct is applied, there are only $p$ contexts need to be considered. (2) Save storage. Obviously, compared with keeping the entire block-traversing history, which requires $L$ storage units for each UE, keeping only two block Ids will significantly reduce the spatial complexity of the system.

#### 4.2.2 Space-Time Contextual BS-Selection Based on Empirical Knowledge of Post-handover Trajectory

Given STCH mechanism and the space-time context of handover, the CC maintains a space-time contextual MAB process for each block to make BS-selection decision for handover events happening in the block. In particular, for any block $g_i$, we define the set of its *preceding blocks* $\mathbf{G}_i = \{g_n^- | g_n = g_i, \forall \tilde{\mathbf{g}}_n = (g_n^-, g_n), \forall n \in \mathbf{U}\}$ from all block concatenations of all UEs, where $\mathbf{U}$ is the set of UEs. The contextual MAB process corresponding to block $g_i$ maintains a coefficient vector $\boldsymbol{\theta}_{i,k}$ for each candidate (or arm) SBS $k \in \mathbf{S}_i$, where $\mathbf{S}_i$ is the set of all available SBSs a UE may handover to in block $g_i$. The elements in $\boldsymbol{\theta}_{i,k}$, each corresponding to a unique preceding block $g_j \in \mathbf{G}_i$, represent the expected rewards an UE in block $g_i$ can receive by handing over to SBS $k$, considering that the UE has come to $g_i$ from various (coarse-grained) directions, respectively (i.e., one element per direction). As described in the previous section, here each coarse-grained direction is represented by an unique preceding block in $\mathbf{G}_i$.

Let $\mathbf{x}$ denote the context which reflects the temporal feature $\tilde{\mathbf{g}}_n = (g_n^-, g_n)$ of a handover event happening on UE $n$ in block $g_i$. Given $\mathbf{x}$ and $\boldsymbol{\theta}_{i,k}$ for $\forall k \in \mathbf{S}_i$, the MAB model will calculate the expected reward for handing over to SBS $k$ as $\bar{r}_{i,k} = \theta^j_{i,k}$, where $\theta^j_{i,k}$ is the element in $\boldsymbol{\theta}_{i,k}$ corresponding to preceding block $g_j$. The actual handover decision is made according to a specific criterion, e.g., choosing the SBS, say $a$, with the largest expected reward. Once a decision is made, the UE will be switched to and served by SBS $a$ from then on. When the LOS connection between the UE and SBS $a$ is lost, e.g., the propagation distance is too long or there is a blockage, an actual reward $r_{i,a}$ that represents the empirical connection time will be calculated. Subsequently, $\mathbf{x}$, $a$, and $r_{i,a}$ will be used to update the expected rewards in $\boldsymbol{\theta}_{i,a}$, as will be described in details in Section 6.

### 4.3 Handover Trigger Condition

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

$$SNR_n^k < SNR_{\min} - h_{\text{ys}}, \tag{6}$$

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

## 5 UCB-Based BS-Selection Algorithm for SCH

The partitioning scheme described in Section 4.1.1 provides the CC the context which reflects the spatial feature of handover and make BS-selection within an specific small Euclidian area, say block. 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 BS-selection in each block as a MAB problem in this section, which is to identify which arm to pull in order to get maximum reward after a given set of trials [44].

### 5.1 Stateless Multi-Armed Bandit Model

Given a block $g_i$ ($i = 1, \ldots, M$, where $M$ is the number of blocks), its corresponding signal vector indicates the candidate SBS set $\mathbf{S}_i$ for the UEs who reside in the block. Let $SNR_i^k$ be the SNR received by a UE from SBS $k$ in block $g_i$, then $\mathbf{S}_i$ is specified as

$$\mathbf{S}_i = \{k \mid SNR_i^k \geq SNR_{\min}, k \in \mathbf{S}\}. \tag{7}$$

After the CC chooses a SBS $k \in \mathbf{S}_i$ for a handover which happens in block $g_i$ in trial $t$ at time $\tau$, the UE will be served by SBS $k$ until it needs another handover, suppose at time $\tau'$. Then the UE receives an instantaneous reward associated with SBS $k$ in block $g_i$, denoted as $r^t_{i,k} = \tau' - \tau$. Since $\tau'$ is an unknown random variable which is determined by the realization of UE's post-handover mobility, including moving trajectory and speed, the reward $r^t_{i,k}$ is also an i.i.d. random variable. As there are no explicit states of SBS as prior knowledge during handover in SCH, the SBS selection in block $g_i$ is formulated by a stateless MAB model [44] $\mathcal{M}_i = \{\mathbf{S}_i, \mu^t_{i,k}\}$, where $k \in \mathbf{S}_i$, and $\mu^t_{i,k}$ is the expected reward of SBS $k$ in block $g_i$ in trial $t$.

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

$$R_{i,T} = \max_{k \in \mathbf{S}_i} \mathbb{E}\left[\sum_{t=1}^{T} r^t_{i,k}\right] - \mathbb{E}\left[\sum_{t=1}^{T} r^t_{i,a_{i,t}}\right]. \tag{8}$$

Based on the model $\mathcal{M}_i$, the handover decision problem in block $g_i$ 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 MAB problem that minimizes the regret.

### 5.2 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 $k^* = \arg\max_{k \in \mathbf{S}_i} \mu^t_{i,k}$ for handover in block $g_i$ all the time. Unfortunately, this assumption does not hold. Therefore, the expected reward of SBS can only be estimated based on historical observations [7]. 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 block $g_i$, respectively. These two metrics are updated by an observation of reward $r^t_{i,k}$ as follows:

$$\bar{r}_{i,k}(T_{i,k}+1) = \frac{T_{i,k} \times \bar{r}_{i,k}(T_{i,k}) + r^t_{i,k}}{T_{i,k}+1}, \tag{9}$$

$$T_{i,k} := T_{i,k} + 1. \tag{10}$$

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 block $g_i$. Each instantaneous reward obtained by any UE is used to update the corresponding mean reward of its serving SBS.

Since the reward is defined as the length of the interval between the moment when the current handover decision is made and the moment when the next handover event happens, it is related to the UE's moving speed. In particular, given two UEs with the same moving trajectory and the same handover decision for them, the one whose speed is low will receive more reward than the other one whose speed is high, because of its long LOS connection time. We consider the random speed of UE as a factor contributing to the randomness of reward of which the distribution can be reflected by the statistic reward accumulated from the past handovers. Therefore, there is no assumption on the distribution of UE's speed in the proposed mechanisms. They can work under any assumption about the distribution of UE's speed in practice. In Section 8, we simulate a random speed scenario in which a UE's moving speed is randomly distributed according to a Gaussian distribution.

### 5.3 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 to avoid being 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 section, we utilize the widely-used UCB policy proposed by [45] to handle this trade-off, since it can achieve logarithmic regret with low computation complexity [7].

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

$$k^* = \arg\max_{k \in \mathbf{S}_i} \left( \bar{r}_{i,k}(T_{i,k}) + \sqrt{\frac{2\ln F_i}{T_{i,k}}} \right). \tag{11}$$

### 5.4 Dynamic Block Set Construction

In SCH, we maintain a MAB model for each block which corresponds to an unique signal vector. However, as mentioned in Section 4.1.1, we are supposed to have no knowledge about the signal space or block partition at the beginning of the algorithm. Hence, we set the block set $\mathbf{G} = \emptyset$ initially. When a handover is triggered, the CC firstly calculates the UE's signal vector. If the signal vector has been observed before, then the Id of the associated block is retrieved; if not, this means that the block where the UE resides has never been identified before. The CC then gives the new identified block an unique Id, suppose to be $g_{\text{new}}$, and adds $g_{\text{new}}$ into $\mathbf{G}$ while keeping the mapping between the new signal vector and the new block Id. Meanwhile, a new MAB process for the new block is created. In this way, the block set is built dynamically.

### 5.5 Acceleration Technique

Generally, when an Event A2 occurs on a LOS connection which was built in block $g_i$ to serve a UE $n$ by SBS $k$, a reward $r_{i,k}$ would be obtained and only $\bar{r}_{i,k}$ would be updated (time- and trial- related subscripts are omitted). However, since the UE's post-handover trajectory is realized at this moment, 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, if the CC switched the UE $n$ to SBS $a$ in block $g_i$, the CC was also aware of the set of SBSs which were not selected, denoted as $\bar{\mathbf{S}}_{i,a} = \mathbf{S}_i \backslash \{a\}$, and pretended to build a virtual LOS link between the UE $n$ and each $k' \in \bar{\mathbf{S}}_{i,a}$. During the UE's post-handover movement, in addition to checking the handover trigger condition on the true LOS link, the CC kept checking that on each virtual LOS link. If the virtual LOS path between the UE $n$ and the SBS $k'$ was blocked, the observed reward $r_{i,k'}$ was 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. In particular, suppose there are average $K_{\text{b}}$ candidate SBSs in a block and SBS $k$ is selected as a handover decision, in the post-handover trajectory of the UE, the CC keeps the virtual connections for the $K_{\text{b}} - 1$ unchosen SBSs and records received virtual rewards. Suppose the CC averagely receives $K_{\text{b}}^{'}$ virtual rewards in the post-handover trajectory, where $K_{\text{b}}^{'} \leq K_{\text{b}} - 1$, then the accumulated rewards $K_{\text{b}}^{'}$ of the unchosen SBSs, besides that of the chosen one, could be simultaneously updated. It means that we are able to use one training sample (i.e., the post-handover trajectory) to update $K_{\text{b}}^{'} + 1$ accumulated rewards, which is supposed to be achieved by using $K_{\text{b}}^{'} + 1$ individual samples if the acceleration technique is not applied. In another word, the learning efficiency is increased by $K_{\text{b}}^{'}$ times by the acceleration technique.

The UCB-based BS-selection algorithm for SCH is summarized in Algorithm 1.

**Algorithm 1:** UCB-based BS-selection algorithm for SCH

**Input:** Cellular network N which consists of a set $\mathbf{S}$ of SBSs and a set of obstacles

$\mathbf{G} = \emptyset$;
**while** *Event A2 handover trigger condition is met for a UE* $n$ **do**
  Record the current time $\tau$;
  Identify the block $g_i$ where UE $n$ resides, associated with the available SBS set $\mathbf{S}_i \subseteq \mathbf{S}$;
  **if** $g_i \notin \mathbf{G}$ **then**
    $T_{i,k} \leftarrow 0$;
    $\bar{r}_{i,k}(0) \leftarrow 0$;
    $F_i \leftarrow 0$;
    $\mathbf{G} \leftarrow \mathbf{G} \cup g_i$;
  $a_i = \arg\max_{k \in \mathbf{S}_i} \left( \bar{r}_{i,k}(T_{i,k}) + \sqrt{\frac{2\ln F_i}{T_{i,k}}} \right)$;
  Switch the UE $n$ to the SBS $a_i$;
  Observe the reward $r_{i,a_i} = \tau' - \tau$ when the next handover occurs for UE $n$ at time $\tau'$;
  $\bar{r}_{i,a_i}(T_{i,k}+1) \leftarrow \frac{T_{i,k} \times \bar{r}_{i,a_i}(T_{i,k}) + r_{i,a_i}}{T_{i,k}+1}$;
  $T_{i,k} \leftarrow T_{i,k} + 1$;
  $F_i \leftarrow F_i + 1$;
  Update $\bar{r}_{i,k'}(T_{i,k'}+1)$, $T_{i,k'}$ and $F_i$ for $k' \in \bar{\mathbf{S}}_{i,a_i}$ in the same way, if the virtual reward $r_{i,k'}$ is obtained;

## 6 LinUCB-Based BS-Selection Algorithm for STCH

According to STCH mechanism described in Section 4.2, the awareness of temporal feature of handover is beneficial for handover decision. In this section, we formulate this decision as a contextual MAB problem and propose a LinUCB-based BS-selection algorithm which considers the time-space feature of handover for STCH.

### 6.1 Contextual Multi-Armed Bandit Model

The basic framework and regret analysis of a contextual bandit algorithm are similar to the algorithm described in Section 5. The main difference between them lies in that,

given the handovers with the same spatial feature, the former further identifies the temporal feature of each handover and applies a personalized policy, while the latter treats them with no difference and applies an identical policy.

The proposed BS-selection algorithm for STCH is based on the LinUCB algorithm which has been widely utilized in many industry fields [46], [47]. In trial $t$ ($t = 1, 2, 3, \ldots, T$), the algorithm observes the handover event to identify the block $g_i$ where it occurs associated with the candidate SBS set $\mathbf{S}_i$. The context (i.e., temporal feature of handover) $\mathbf{x}_{i,k}^t$ with dimension of $d_i$, which is associated with SBS $k \in \mathbf{S}_i$, is then extracted, and the details will be discussed in Section 6.2. According to [46], for all $t$, the expected reward received from SBS $k$ is linear in its $d_i$-dimension context $\mathbf{x}_{i,k}^t$ with unknown coefficient vector $\boldsymbol{\theta}_{i,k}$ and is shown as

$$\bar{r}_{i,k}^t = \mathbb{E}[r_{i,k}^t | \mathbf{x}_{i,k}^t] = {\mathbf{x}_{i,k}^t}^\top \boldsymbol{\theta}_{i,k}. \tag{12}$$

The $d_i$-dimension coefficient vectors $\boldsymbol{\theta}_{i,k}$ are adapted based on the accumulated observations and used to guide the future handover decision accompanied with the context $\mathbf{x}_{i,k}^t$. Denote $\mathbf{D}_{i,k}$ as a $m \times d_i$ matrix which consists of $m$ contexts observed previously for SBS $k$ in block $g_i$, and $\mathbf{c}_{i,k}$ as a $m$-dimension vector which indicates the rewards of the $m$ observations. The optimal $\boldsymbol{\theta}_{i,k}^*$ is estimated by applying ridge regression to $\mathbf{D}_{i,k}$ and $\mathbf{c}_{i,k}$:

$$\hat{\boldsymbol{\theta}}_{i,k} = (\mathbf{D}_{i,k}^\top \mathbf{D}_{i,k} + \mathbf{I}_{d_i})^{-1} \mathbf{D}_{i,k}^\top \mathbf{c}_{i,k}, \tag{13}$$

where $\mathbf{I}_{d_i}$ is a $d_i$-dimension identity matrix. In each trial $t$ in block $g_i$, the algorithm chooses SBS

$$a_{i,t} = \arg\max_{k \in \mathbf{S}_i} \left( {\mathbf{x}_{i,k}^t}^\top \hat{\boldsymbol{\theta}}_{i,k} + \eta \sqrt{{\mathbf{x}_{i,k}^t}^\top \mathbf{A}_{i,k}^{-1} \mathbf{x}_{i,k}^t} \right) \tag{14}$$

as the handover result, where $\mathbf{A}_{i,k} = \mathbf{D}_{i,k}^\top \mathbf{D}_{i,k} + \mathbf{I}_{d_i}$, and $\eta > 0$ is the hyper-parameter. Specifically, the first item in Eq. (14) is the predicted reward for SBS $k$, while $\sqrt{{\mathbf{x}_{i,k}^t}^\top \mathbf{A}_{i,k}^{-1} \mathbf{x}_{i,k}^t}$ indicates the standard deviation of reward. Given previous observations, the algorithm chooses SBS $a_{i,t}$ with the optimal expected reward according to $\mathbf{x}_{i,a_{i,t}}^t$. When the reward $r_{i,a_{i,t}}^t$ is obtained, the new observation $(\mathbf{x}_{i,a_{i,t}}^t, a_{i,t}, r_{t,a_{i,t}}^t)$ is used to improve the BS-selection policy. The goal of the algorithm is also to find the optimal policy to minimize the regret which is formulated as Eq. (8). It has been proven that LinUCB algorithm has good ability to implement exploitation-exploration trade-off [46].

### 6.2 Context Construction

The design of context for a contextual MAB model depends on the characteristics of the specific problem. In our handover problem, once there is a handover triggered, the CC firstly identifies its spatial feature, i.e., the block where it occurs, and then extract its temporal feature based on the UE's block concatenation. These features are then used to construct the context of the proposed contextual MAB model for STCH.

In particular, given the preceding blocks $\mathbf{G}_i$, defined in Section 4.2.2, of any block $g_i$ and $d_i = |\mathbf{G}_i|$, $i = 1, \ldots, M$, we define the context associated with SBS $k \in \mathbf{S}_i$ for any handover event requested by UE $n$ in block $g_i$ in trial $t$ is defined as a $d_i$-dimension 0-1 vector, denoted by $\mathbf{x}_{i,k}^t = (x_{i1}^t, \ldots, x_{i,d_i}^t)$. Each element $x_{ij}^t \in \mathbf{x}_{i,k}^t$ corresponds to a proceeding block $g_{ij} \in \mathbf{G}_i$, where $g_{ij}$ denotes the $j$th preceding block in $\mathbf{G}_i$. Note that we consider that all SBS $k \in \mathbf{S}_i$ share the same context construction. The value of $x_{ij}^t \in \mathbf{x}_{i,k}^t$ depends on the temporal feature of the handover, say $\tilde{\mathbf{g}}_n = (g_n^-, g_n)$, which is indicated by the UE's block concatenation, where $g_n = g_i$. Specifically, $x_{ij}^t = 1$ if $g_{ij} = g_n^-$, and $x_{ij}^t = 0$ otherwise, where $j = 1, \ldots, d_i$.

Similar to SCH, the block set in STCH also needs to be dynamically constructed during the algorithm, as described in Section 5.4. The difference is, in SCH, the block identifying is conducted only when a handover is triggered, while in STCH, it is conducted all the time during UE's movement in order to keep updating UE's block concatenation. Note that, the set of preceding blocks $\mathbf{G}_i$ of any block $g_i$ is also dynamically maintained. Due to the lack of prior knowledge, initially we set $\mathbf{G}_i = \emptyset$, $\forall g_i \in \mathbf{G}$. During the procedure of the algorithm, when given a preceding block $g_n^-$ associated with a handover event occurring in block $g_i$, we add it into $\mathbf{G}_i$ if $g_n^- \notin \mathbf{G}_i$. Meanwhile, the dimension of the temporal feature of any handover event occurring in the block $g_i$ is increased by 1, and the dimensions of related coefficients are expended accordingly.

### 6.3 Acceleration Technique

The acceleration technique introduced in Section 5.5 can also be applied to improve the efficiency of the LinUCB-based BS-selection algorithm for STCH. Specifically, when handing over a UE $n$ to the chosen SBS, suppose to be SBS $k$, in block $g_i$ in trial $t$ with context $\mathbf{x}_{i,k}^t$, the CC also build a virtual connection from the UE to each SBS in $\bar{\mathbf{S}}_{i,k}$. In the UE's post-handover trajectory, if a virtual connection between the UE to a SBS $k' \in \bar{\mathbf{S}}_{i,k}$, is blocked, the instantaneous virtual reward $r_{i,k'}^t$ is obtained. Then the associated virtual observation $(\mathbf{x}_{i,k'}^t, k', r_{i,k'}^t)$ is used to update the coefficients for SBS $k'$ in the LinUCB model of block $g_i$. In this way, not only the coefficients for SBS $k$, but also those for some SBS $k' \in \bar{\mathbf{S}}_{i,k}$ can be updated by a single realization of UE's post-handover trajectory. As mentioned in Section 5.5, the virtual update can fully exploit true experience to increase the efficiency of the BS-selection algorithm.

The LinUCB-based BS-selection algorithm for STCH is summarized in Algorithm 2.

## 7 Complexity Analysis

The computation complexity includes spatial complexity and temporal complexity. By complexity analysis, the total costs of the system with the proposed algorithms consist of the following three parts:

(1) MAB related cost. The CC maintains a MAB process for each block. For an UCB-based BS-selection algorithm, the spatial complexity is $\mathbf{O}(K_\mathrm{b})$, where $K_\mathrm{b}$ is the average number of candidate SBSs in a block. Specifically, each arm requires two storage units to maintain its expected reward and its chosen times, respectively. Moreover, the temporal complexity is also $\mathbf{O}(K_\mathrm{b})$, since the CC has to compute and search for the biggest index among the arms, i.e., the candidate SBSs. Similarly, for a LinUCB-based BS-selection

algorithm, the temporal complexity is also $\mathbf{O}(K_\mathrm{b})$. Due to the time-space context considered in STCH, extra storage units are required to keep the feature-related information for each arm, and the spatial complexity is $\mathbf{O}(d_\mathrm{b}K_\mathrm{b})$, where $d_\mathrm{b}$ is the average dimension of the temporal feature of a handover in a block. For the whole system, there are $M$ MAB processes, where $M$ is the number of the blocks as well as the signal vectors identified according to the signal space partitioning scheme. Theoretically, the amount of blocks are huge. Suppose there are $K$ SBSs in the network and $J$ quantizing thresholds, there exist totally $(J+1)^K$ different signal vectors in the signal space. However, because the propagation range of mmWave signal is limited, the number of the available SBSs for a UE is much smaller than that of the whole SBSs. It means that the signal space is sparse and the vast majority of the signal vectors will not show up in practice. Therefore, the number of MAB processes $M \ll (J+1)^K$. Hence, the total MAB related spatial complexity is $\mathbf{O}(MK_\mathrm{b})$ for SCH, and $\mathbf{O}(Md_\mathrm{b}K_\mathrm{b})$ for STCH.

**Algorithm 2:** LinUCB-based BS-selection algorithm for STCH

**Input:** Cellular network N which consists of a set $\mathbf{S}$ of SBSs and a set of obstacles, $\eta \in \mathbb{R}_+$

$\mathbf{G} = \emptyset$;
Keep tracking each UE $n$ and updating its block concatenation $\tilde{\mathbf{g}}_n = (g_n^-, g_n)$;
**if** $g_n \notin \mathbf{G}$ **then**
  $\mathbf{G}_n = \emptyset$, $d_n = 0$;
  $\mathbf{A}_{n,k} = \emptyset$, $\mathbf{b}_{n,k} = \emptyset$;
  $\mathbf{G} \leftarrow \mathbf{G} \cup g_n$;
**while** *Event A2 handover trigger condition is met for UE* $n$ **do**
  Identify the block $g_i$ where UE $n$ resides, associated with the available SBS set $\mathbf{S}_i$ and the corresponding trial number $t$;
  Retrieve $\tilde{\mathbf{g}}_n = (g_n^-, g_i)$ for UE $n$ from the memory;
  **if** $g_n^- \notin \mathbf{G}_i$ **then**
    $\mathbf{A}_{i,k} \leftarrow [\mathbf{A}_{i,k}|v_0]$, where $v_0$ is a $d_i$-dimension zero vector;
    $\mathbf{A}_{i,k}^\top \leftarrow [\mathbf{A}_{i,k}^\top|v_1]$, where $v_1$ is a $(d_i+1)$-dimension vector, $v_1 = [0, \ldots, 0, 1]$;
    $\mathbf{b}_{i,k} \leftarrow [\mathbf{b}_{i,k}|0]$;
    $d_i \leftarrow d_i + 1$;
    $\mathbf{G}_i \leftarrow \mathbf{G}_i \cup g_n^-$;
  Observe context $\mathbf{x}_{i,k}^t$ from $\tilde{\mathbf{g}}_n$ for all SBS $k \in \mathbf{S}_i$;
  $\hat{\boldsymbol{\theta}}_{i,k} \leftarrow \mathbf{A}_{i,k}^{-1}\mathbf{b}_{i,k}$;
  $q_{i,k}^t \leftarrow {\mathbf{x}_{i,k}^t}^\top \hat{\boldsymbol{\theta}}_{i,k} + \eta\sqrt{{\mathbf{x}_{i,k}^t}^\top \mathbf{A}_{i,k}^{-1}\mathbf{x}_{i,k}^t}$;
  Choose SBS $a_{i,t} = \arg\max_{k \in \mathbf{S}_i} q_{i,k}^t$ and observe a reward $r_{i,k}^t$;
  $\mathbf{A}_{i,a_{i,t}} \leftarrow \mathbf{A}_{i,a_{i,t}} + \mathbf{x}_{i,a_{i,t}}^t {\mathbf{x}_{i,k_{i,t}}^t}^\top$;
  $\mathbf{b}_{i,a_{i,t}} \leftarrow \mathbf{b}_{i,a_{i,t}} + r_{i,k}^t \mathbf{x}_{i,a_{i,t}}^t$;
  Update $\mathbf{A}_{i,k'}$ and $\mathbf{b}_{i,k'}$ for $k' \in \bar{\mathbf{S}}_{i,a_{i,t}}$ in the same way, if the virtual reward $r_{i,k'}^t$ is obtained;

TABLE 2
Complexity Analysis

| | Temporal Complexity | Spatial Complexity |
|---|---|---|
| SCH | $\mathbf{O}(M + K_\mathrm{b})$ | $\mathbf{O}(MK_\mathrm{b} + U)$ |
| SCH with acceleration | $\mathbf{O}(M + K_\mathrm{b})$ | $\mathbf{O}(MK_\mathrm{b} + UK_\mathrm{b})$ |
| STCH | $\mathbf{O}(M + K_\mathrm{b})$ | $\mathbf{O}(Md_\mathrm{b}K_\mathrm{b} + UL)$ |
| STCH with acceleration | $\mathbf{O}(M + K_\mathrm{b})$ | $\mathbf{O}(Md_\mathrm{b}K_\mathrm{b} + U(L + K_\mathrm{b}))$ |

(2) Block set related cost. Since the CC keeps the mapping between each block and its associated signal vector, the spatial complexity to maintain the signal vector and the blocks is $\mathbf{O}(M)$. When a handover is triggered, the CC needs to retrieve the Id of the block where the handover event occurs according to a given signal vector. This search contributes the temporal complexity of $\mathbf{O}(M)$.

(3) UE related cost. In order to support the BS-selection algorithm, the CC needs to keep the reward information for all UEs, and the spatial complexity is $\mathbf{O}(U)$, where $U$ is the average number of all UEs in the network over time. If the acceleration technique, i.e., virtual update, is applied, the CC also needs to keep checking the virtual LOS links and the extra spatial complexity of $\mathbf{O}(UK_\mathrm{b})$ is required to keep the virtual rewards for all UEs. Especially, for STCH, the CC also needs to keep a block concatenation for each UE, which requires additional spatial complexity of $\mathbf{O}(LU)$, where $L$ is the size of the block concatenation. When the one-step-back-look construct is used, $L$ could be reduced to 2.

To sum up, the total costs as well as the computation complexity of the proposed schemes could be summarized in Table 2.

The proposed algorithms only require additional costs at the CC side and there is no much cost at the UE side. The extra power consumption at the UE is to report RSS information to the CC, but this power consumption is minimal because of the small amount of data involved in the report. Because the proposed MAB algorithm is an online reinforcement learning algorithm, it does not require an offline training before the algorithm can be deployed online and start functioning. Instead, the algorithm can be operational online at time 0. Moreover, since these schemes only add temporal complexity which is linear in the number of candidate SBSs, the impact on the latency is limited.

## 8 Numerical Simulations

In this section, we evaluate the performances of two proposed contextual handover mechanisms, SCH and STCH, in various scenarios. Theoretically, the accumulated regret is the common metric to evaluate a solution of a MAB problem. However, in our problem, the optimal handover decision in hindsight is difficult to compute due to the large scale of the problem. Therefore, instead of comparing directly with the ground-truth optimal solution, we take an indirect but practical perspective and compare our proposed (approximate) algorithms with other counterpart (approximate) schemes, i.e., Rate-first hanvover (RFH) and SMART, over two new metrics, i.e., the average number of handovers per UE (ANH) and the average lasting time per each LOS

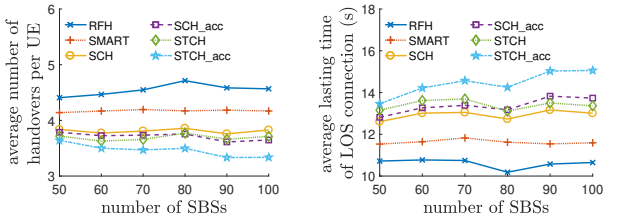
connection (ALT), to gauge our achievable performance gains over those schemes that are actually implementable in practice. Specifically, in RFH, the BS that provides the maximum transmission rate calculated by Equation (4) is chosen as the handover result, while SMART proposes a reinforcement-learning based BS-selection algorithm to make handover decisions. We believe these two metrics, ANH and ALT, can well reflect the average QoE of an arbitrary user.

### 8.1 Simulation Settings

We consider a cellular network N that covers a $100(\mathrm{m}) \times 100(\mathrm{m})$ square region and consists of a certain number of SBSs (100 by default) using mmWave band. In this setting, we calculate pathloss, SNR and throughput according to Eqs. (1)-(4). The transmit power of SBS is set to be 30 dBm, and the noise power is -57 dBm. Similar to [35], we set the parameters $\alpha$ and $\beta$ in Eq. (1) as 61.4 and 2, respectively, corresponding to a carrier frequency of 28 GHz. The channel gain $G_{\max}$ of main lobe is 18 dB as in [36]. The bandwidth of SBS is set as 1000 MHz, and $U_{\max}$ is set to be 10. We assume that a certain number of identical obstacles (20 by default) with radius of 1(m) are randomly distributed in the network. The minimum required transmission rate $h_{\min}$ in (6) is set to be 1000 Mbps. The number of UEs entering into the network per time slot follows a Poisson distribution with parameter $\lambda$. For a new coming UE, its initial position is uniformly distributed at the border of the network and its moving direction is also uniformly distributed. The UE's moving speed is supposed to be random following $\mathcal{N}(\mu_{\mathrm{v}}, \sigma_{\mathrm{v}}^2)$, where $\mu_{\mathrm{v}}$ is the mean value and $\sigma_{\mathrm{v}}$ is the standard deviation, and are set to be 1 and 0.1 by default, respectively. Any UE's experience is used to update the accumulated reward received from the serving SBS until it moves out of the network region. The hyper-parameter $\eta$ in Eq. (14) is set to be 0.25 empirically because this value leads to the best performance within our simulations. The other key parameters, i.e., the number of SBSs, the number of obstacles, the UE's arrival rate and the threshold of SNR, are set to be 100, 20, 3 (/iteration) and 20 (dB), respectively by default, if not specified.

### 8.2 Density of SBSs

In this simulation, we compare the performances of the candidate handover mechanisms, i.e., RFH, SMART, SCH, SCH with acceleration (SCH_acc), STCH and STCH with acceleration (STCH_acc), with different numbers of SBSs on ANH and ALT. We choose six values for the number of SBS: 50, 60, 70, 80, 90 and 100, and run 10000 iterations (time unit) for each instance. The results are shown in Fig. 4. It can be found that, although SMART outperforms RFH on ANH and ALT by 11.5% and 14.1%, respectively, the proposed contextual handover mechanisms perform better than both of them. Compared with SMART, SCH improves ANH and ALT by up to 10.0% and 14.0%, while STCH improves even better, by up to 12.4% and 17.0%, respectively when given 90 SBSs. This means, by considering temporal feature, we can obtain better handover decision than by only taking account of spatial feature. When acceleration technique is applied, still compared with SMART, SCH_acc improves ANH and ALT by 13.6% and 19.8%, while STCH_acc improves by 20.3% and 30.2% , respectively. This demonstrates the efficiency improved by the acceleration technique. Note that, when the number of SBSs grows, there is no significant improvement on the performances of SCH and STCH. This is because that when there are more SBSs, there will be more blocks identified and more MAB processes maintained according to the signal space partitioning scheme. Therefore, with given iterations, the average number of training samples obtained by each MAB model is smaller and the models easily become undertrained. Fortunately, we can address this issue by applying the acceleration technique which increases the utility efficiency of each training sample, and achieve good performance even with limited iterations.

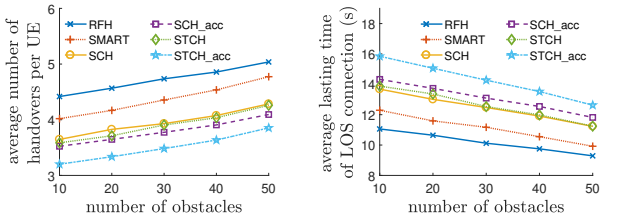



(a) Performance on the average number of handovers per UE (b) Performance on the average lasting time per LOS connection

Fig. 4. Comparison of performances with different numbers of SBSs

### 8.3 Density of Obstacles

Obstacle on LOS path is one of the main reasons which render frequent outages in mmWave communication. In this simulation, we investigate the impact of the density of obstacles on the performances of the candidate handover mechanisms. Fig. 5 illustrates ANH and ALT obtained under these mechanisms with various numbers of obstacles, from 10 to 50, after 10000 iterations. With the growth of the number of obstacles, more handovers occur and the lasting time of LOS connection decrease, no matter under which mechanism. This trend is accord with the intuition that, there will be more handover events in a complex environment with many obstacles than in a simple environment with few obstacles. Furthermore, the proposed handover mechanisms, i.e., SCH, SCH_acc, STCH and STCH_acc, observably outperform the other two, i.e., RFH and SMART, no matter in a complex or a simple environment.



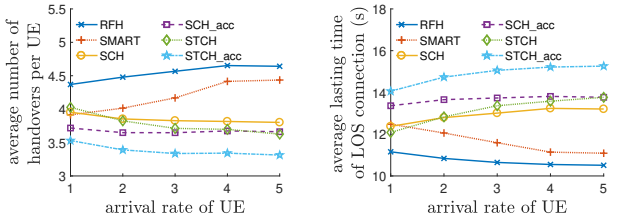

(a) Performance on the average number of handovers per UE (b) Performance on the average lasting time per LOS connection

Fig. 5. Comparison of performances with different numbers of obstacles

### 8.4 UE's Arrival Rate

In this simulation, we compare the performances of these candidate handover mechanisms 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 $\lambda$: 1, 2, 3, 4 and 5, and run 10000 iterations for each instance. The results are displayed in Fig. 6. As shown, in a crowded scenario (i.e., when $\lambda$ approaches to 5), the performances of RFH and SMART decrease, while the proposed mechanisms all have good performances, which are even better than those in an uncrowded scenario. This result shows that the proposed mechanisms have good performance even in a crowded scenario. The reason is that, in an crowded network, there are so many users to provide enough training samples to make the MAB models well trained. While in an uncrowded situation, the MAB models may get undertrained since there are not enough training samples. This is the reason why two out of the four proposed mechanisms, i.e., SCH and STCH, perform even worse than SMART when $\lambda = 1$. However, if well trained in a crowded scenario, their performances could be improved significantly. Specially, the performance of STCH are even better than that of SCH_acc when $\lambda = 5$. This is another demonstration of the advantage of space-time context over spatial context.




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

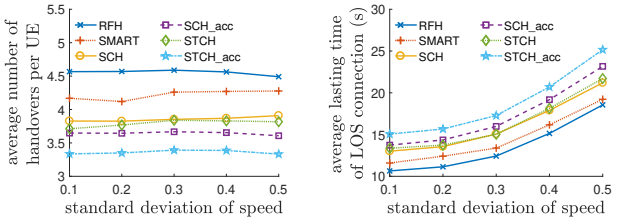
Fig. 6. Comparison of performances with different arrival rates

### 8.5 Variance of UE's Speed

As described in Section 4.2, we consider random moving speed of UEs following Gaussian distribution, denoted as $\mathcal{N}(1, \sigma_v^2)$, in our simulations. Since the instantaneous reward obtained by a UE is impacted by its specific moving speed, we investigate the performances of the proposed mechanisms with different variances of UE's moving speed in this simulation. As the variance of UE's speed is indicated by $\sigma_v$, we choose five values for $\sigma_v$, from 0.1 to 0.5, and run 10000 iterations for each instance. The simulation results are shown in Fig. 7. As expected, the performances of the proposed mechanisms on ANH and ALT outperform the two benchmarks. Note that, the ANHs under all candidate mechanisms generally keep the same, no matter with which variance of speed. The reason lies in that, given the distribution of speed, the impact of the variance of reward caused by random speeds of UE on the performance of ANH could be averaged out over time. This result demonstrates the stability of the proposed handover mechanisms in the scenario with UEs whose speeds are various.

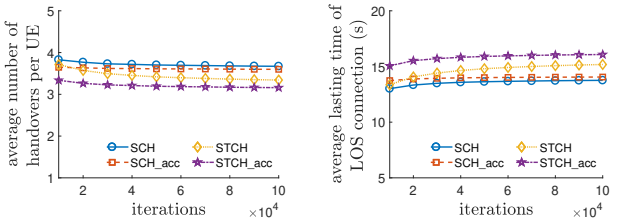
### 8.6 Convergence Evaluation

To evaluate the convergence of the BS-selection algorithms and the acceleration technique in the proposed mechanisms, we adopt the default settings and run these mechanisms for 100000 iterations in this simulation. Here, we do not take RFH and SMART into consideration. The simulation results are shown in Fig. 8. It is obvious that, although the performances of all mechanisms approach to convergence when the algorithms stop, their behaviors are not the same. Specifically, compared with SCH, SCH_acc shows better convergence due to the acceleration technique. The same result can be obtained by comparing STCH and STCH_acc. Although the performance of STCH outperforms those of SCH and SCH_acc eventually, it needs longer time to get converged. It is because the training samples obtained in a block are divided by the temporal features of handover and hence more samples are needed to have the MAB process well trained. Fortunately, this backward can be overcome by the acceleration technique, as demonstrated by the curve of STCH_acc which presents good convergence even within limited iterations.




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

Fig. 7. Comparison of performances with different variances of UE's speed



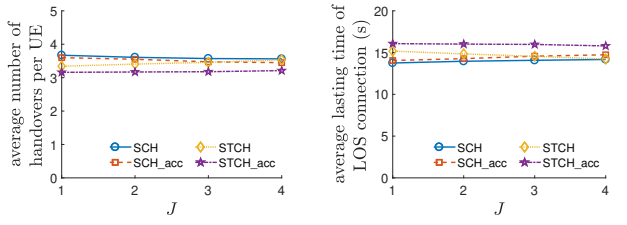

(a) Performance on the average number of handovers per UE (b) Performance on the average lasting time per LOS connection

Fig. 8. Comparison of performances with different iterations

### 8.7 Granularity of Partition

As we described in Section 4.1.1, the quantizing thresholds play an important role to determine the granularity of signal space partition. Theoretically, finer granularity of partitioning would lead to better handover decisions. In this simulation, we investigate the impact of the number of quantizing thresholds on the performances of the proposed mechanisms. We try four values for $J$: 1, 2, 3 and 4, and run 100000 iterations for each instance. The results are displayed in Fig. 9. Similar to Section 8.6, we do not take account RFH and SMART in this simulation, since they do not refer to the signal space partitioning scheme. The simulation results show that, for most mechanisms, although the performances under fine granularity (i.e., $J = 5$) of partitioning are better than those under coarse granularity (i.e., $J = 1$), their difference is ignorable. However, the increment of required storage with fine granularity is much larger than that with coarse granularity. We show the number of identified blocks under the scenarios with different values of $J$ in Fig. 10. As shown, with the growth of $J$, the number of identified blocks increases exponentially. That means the

CC will spend much higher storage cost to support the fine granularity of partitioning. Obviously, the input is not proportional to the output when we adopt fine granularity of partitioning. In another word, coarse granularity of partitioning, i.e., $J = 1$, is good enough to achieve good performance while requiring the lowest computation cost.




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

Fig. 9. Comparison of performances with different numbers of quantizing thresholds

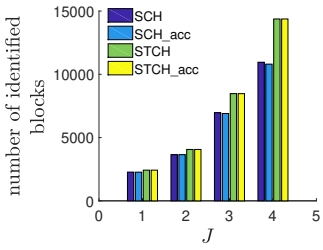



Fig. 10. Increase of the number of identified blocks

### 8.8 Regulation of UE's Movement

In the above simulations, the UE's initial position and moving direction are set to be 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 UEs normally move along with the existing sidewalks, the UE's position is actually limited within the area of a sidewalk but not the whole network, and the UE's moving direction is along with the sidewalk. In this simulation, we investigate the impact of regulation of UE's movement on the performances of the candidate handover mechanisms. We adopt the grid-based scenario according to [48], [49] and divide the whole network into 400 identical square areas each with the size of 5(m) $\times$ 5(m). Note that, these square areas are used to deploy SBSs regularly and have nothing to do with the blocks we mentioned before. In this grid-based network, we set four sidewalks which are indicated by grey square areas where the UEs' moving directions are along with these sidewalks. In particular, the UEs are generated at one end of a sidewalk and their moving directions are indicated by the corresponding arrows. We deploy 80 SBSs in the network, represented by blue circles, and each of them locates at the center of a square area along the sidewalks. This deployment simulates the sidewalks with street lamps equipped with SBSs. Besides, 20 obstacles are randomly distributed in the sidewalks, represented by red triangles.

In order to express different degrees of regulation for UE's movement, we introduce a specific parameter $\gamma \in [0.1, 0.5]$ to describe the homogeneity of UEs' movements within the sidewalks. In particular, if $\gamma = 0.1$, ten percents of UEs in the same sidewalk move with the same direction

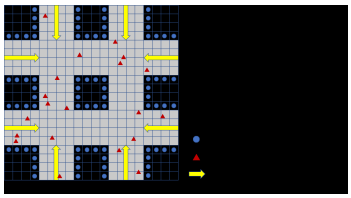


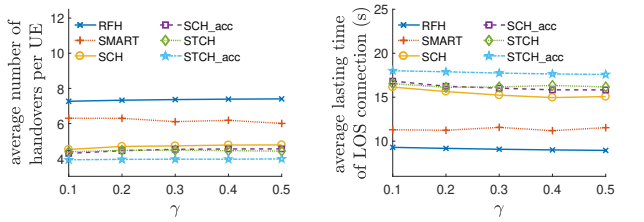
Fig. 11. Scenario with sidewalks

while ninety percents move oppositely; if $\gamma = 0.5$, half of the UEs in the same sidewalk move oppositely. Besides, we choose different values for $SNR_{\min}$: 18, 19 and 20, and run 10000 iterations for each instance with distinct combination of $\gamma$ and $SNR_{\min}$. The simulation results are shown in Fig. 12. Not surprising, the proposed contextual handover mechanisms are superior to the other two, no matter in the scenario with regular movement or in the scenario with irregular movement of UE. More important, the performance of the proposed mechanisms does not change significantly when UEs' movement become irregular. That means our mechanisms have stable performances in a general situation in real-world.

Moreover, we are also interested with the performance difference between the proposed mechanisms with different values of $SNR_{\min}$. Comparing Fig. 12(a), 12(c) and 12(e), it is obvious that, the performances of the proposed mechanisms when $SNR_{\min} = 18$ is better than those when $SNR_{\min} = 20$. It is because that, for a handover event, the candidate SBSs with a lower $SNR_{\min}$ are more than those with a higher requirement. With more candidate SBSs, better handover decisions would be made. The same conclusion could be made by comparing Fig. 12(b), 12(d) and 12(f). In addition, the performance differences among the proposed mechanisms become smaller when $SNR_{\min}$ increases. Specifically, when $SNR_{\min} = 18$, the performance differences between the best mechanism, i.e., STCH_acc, and the worst one, i.e., SCH, is 16.6% and 16.7% on ANH and ALT, respectively. However, when $SNR_{\min}$ increases to 20, these two differences significantly reduce to 5.0% and 4.3%, respectively. That means, the advantage of temporal context could be fully exploited with a low $SNR_{\min}$. In another word, with a high $SNR_{\min}$, there is no big difference between the spatial context and the temporal context.

## 9 Conclusions and Future Work

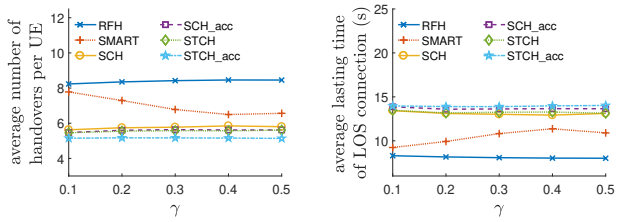
In this paper, we have demonstrated the significant benefits of exploiting UE's spatial and temporal contexts, even a coarse-grained one, in making better handover decisions in ultra-dense mmWave cellular networks. In particular, we have casted the handover decision making as a MAB problem that attempts to minimize the expected regret between the handover in question and the optimal handover decision making in hindsight by learning from the spatial and temporal features of past handovers, respectively. All contexts exploited in the proposed framework are of a coarse-grained nature: the spatial feature of a UE's handover is represented by the coarse-grained area where the handover happened, while its temporal feature is defined as the coarse-grained moving direction of the UE, modeled by the sequence of past areas that the UE has traversed, wherein the areas are defined by a partition of the signal space based on the RSS from nearby SBSs. So




(a) Performance on the average number of handovers per UE ($SNR_{\min} = 18$)



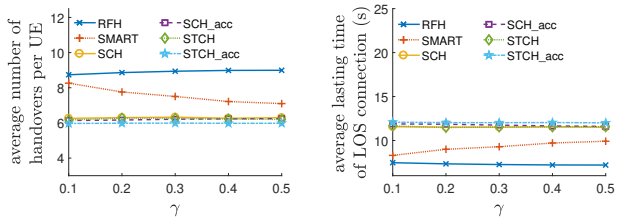

(b) Performance on the average lasting time per LOS connection ($SNR_{\min} = 18$)




(c) Performance on the average number of handovers per UE ($SNR_{\min} = 19$)




(d) Performance on the average lasting time per LOS connection ($SNR_{\min} = 19$)




(e) Performance on the average number of handovers per UE ($SNR_{\min} = 20$)




(f) Performance on the average lasting time per LOS connection ($SNR_{\min} = 20$)

Fig. 12. Comparison of performances on different moving directions of UE

the proposed framework does not require any UE's fine-grained information such as its exact location. A UCB-based solution and a LinUCB-based solution have been proposed to solve the above two MAB formulations, respectively. Our simulation results have shown that by exploiting the coarse-grained spatial contextual information, the proposed UCB-based handover mechanism outperforms existing counterparts that do not utilize this information. Furthermore, by taking advantage of the coarse-grained moving direction information, the LinUCB-based mechanism achieves even better performance than the UCB-based, even for a very simple one-step-look-back implementation.

Our work may be improved from the following directions. First, as a network-layer issue, our proposed handover methods do not consider any physical-layer details, e.g., beamforming. However, it is definitely possible to achieve even better performance improvement if a cross-layer optimization framework is taken, which considers the PHY and link layer features when optimizing a network-layer handover decision. Such a cross-layer optimization is worth studying in our future work. In addition, since this work only considers static obstacles, we will study the case of mobile obstacles in our future research.

## 10 Acknowledgement

This work 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 *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, IEEE, 2014.

[2] B. Van Quang, R. V. Prasad, and I. Niemegeers, "A survey on handoffs—lessons for 60 ghz based wireless systems," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 1, pp. 64–86, 2010.

[3] M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Multi-connectivity in 5g mmwave cellular networks," in *2016 Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, pp. 1–7, IEEE, 2016.

[4] F. Zhao, H. Tian, G. Nie, and H. Wu, "Received signal strength prediction based multi-connectivity handover scheme for ultra-dense networks," in *2018 24th Asia-Pacific Conference on Communications (APCC)*, pp. 233–238, IEEE, 2018.

[5] C. Chaieb, Z. Mlika, F. Abdelkefi, and W. Ajib, "Mobility-aware user association in hetnets with millimeter wave base stations," in *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pp. 153–157, IEEE, 2018.

[6] M. Mezzavilla, S. Goyal, S. Panwar, S. Rangan, and M. Zorzi, "An mdp model for optimal handover decisions in mmwave cellular networks," in *2016 European conference on networks and communications (EuCNC)*, pp. 100–105, IEEE, 2016.

[7] 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 Transactions on Mobile Computing*, no. 6, pp. 1456–1468, 2018.

[8] Q. Li, M. Ding, C. Ma, C. Liu, Z. Lin, and Y.-C. Liang, "A reinforcement learning based user association algorithm for uav networks," in *2018 28th International Telecommunication Networks and Applications Conference (ITNAC)*, pp. 1–6, IEEE, 2018.

[9] Z. Wang, L. Li, Y. Xu, H. Tian, and S. Cui, "Handover control in wireless systems via asynchronous multiuser deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4296–4307, 2018.

[10] Z. Han, T. Lei, Z. Lu, X. Wen, W. Zheng, and L. Guo, "Artificial intelligence-based handoff management for dense wlans: A deep reinforcement learning approach," *IEEE Access*, vol. 7, pp. 31688–31701, 2019.

[11] J. Bao, T. Shu, and H. Li, "Handover prediction based on geometry method in mmwave communications-a sensing approach," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, IEEE, 2018.

[12] 3GPP, *Study on RAN sharing enhancements on GERAN and UTRAN*. 3GPP Release 13. Available [online]: http://www.3gpp.org/dynareport/ganttchart-level-2.htm.

[13] I. Colin, A. Thomas, and M. Draief, "Parallel contextual bandits in wireless handover optimization," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 258–265, IEEE, 2018.

[14] 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.

[15] R. Parada and M. Zorzi, "Context-aware handover in mmwave 5g using ue's direction of pass," in *European Wireless 2018; 24th European Wireless Conference*, pp. 1–6, VDE, 2018.

[16] L. Sun, J. Hou, and T. Shu, "Optimal handover policy for mmwave cellular networks: A multi-armed bandit approach," in *GLOBECOM 2019-2019 IEEE Global Communications Conference*, IEEE, to appear.

[17] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave cellular wireless networks: Potentials and challenges," *arXiv preprint arXiv:1401.2560*, 2014.

[18] F. B. Tesema, A. Awada, I. Viering, M. Simsek, and G. P. Fettweis, "Mobility modeling and performance evaluation of multi-connectivity in 5g intra-frequency networks," in *2015 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, IEEE, 2015.

[19] 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.

[20] Y. Zhu, Z. Zhang, Z. Marzi, C. Nelson, U. Madhow, B. Y. Zhao, and H. Zheng, "Demystifying 60ghz outdoor picocells," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, pp. 5–16, ACM, 2014.

[21] G. Athanasiou, P. C. Weeraddana, C. Fischione, and L. Tassiulas, "Optimizing client association for load balancing and fairness in millimeter-wave wireless networks," *IEEE/ACM Transactions on Networking*, no. 3, pp. 836–850, 2015.

[22] Y. Xu, H. Shokri-Ghadikolaei, and C. Fischione, "Distributed association and relaying with fairness in millimeter wave networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 7955–7970, 2016.

[23] M. Giordani, M. Mezzavilla, A. Dhananjay, S. Rangan, and M. Zorzi, "Channel dynamics and snr tracking in millimeter wave cellular systems," in *European Wireless 2016; 22th European Wireless Conference*, pp. 1–8, VDE, 2016.

[24] S. Goyal, M. Mezzavilla, S. Rangan, S. Panwar, and M. Zorzi, "User association in 5g mmwave networks," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, IEEE, 2017.

[25] C. N. Barati, S. A. Hosseini, M. Mezzavilla, T. Korakis, S. S. Panwar, S. Rangan, and M. Zorzi, "Initial access in millimeter wave cellular systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 7926–7940, 2016.

[26] R. Ford, S. Rangan, E. Mellios, D. Kong, and A. Nix, "Markov channel-based performance analysis for millimeter wave mobile networks," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, IEEE, 2017.

[27] S. Zang, W. Bao, P. L. Yeoh, H. Chen, Z. Lin, B. Vucetic, and Y. Li, "Mobility handover optimization in millimeter wave heterogeneous networks," in *2017 17th International symposium on communications and information technologies (ISCIT)*, pp. 1–6, IEEE, 2017.

[28] J. Qiao, Y. He, and X. S. Shen, "Proactive caching for mobile video streaming in millimeter wave 5g networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 10, pp. 7187–7198, 2016.

[29] O. Semiari, W. Saad, M. Bennis, and B. Maham, "Mobility management for heterogeneous networks: Leveraging millimeter wave for seamless handover," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2017.

[30] O. Semiari, W. Saad, M. Bennis, and B. Maham, "Caching meets millimeter wave communications for enhanced mobility management in 5g networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 779–793, 2018.

[31] F. B. Mismar and B. L. Evans, "Partially blind handovers for mmwave new radio aided by sub-6 ghz lte signaling," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–5, IEEE, 2018.

[32] A. Alkhateeb, I. Beltagy, and S. Alex, "Machine learning for reliable mmwave systems: Blockage prediction and proactive handoff," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1055–1059, IEEE, 2018.

[33] T. Nishio, H. Okamoto, K. Nakashima, Y. Koda, K. Yamamoto, M. Morikura, Y. Asai, and R. Miyatake, "Proactive received power prediction using machine learning and depth images for mmwave networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 11, pp. 2413–2427, 2019.

[34] Y. Koda, K. Yamamoto, T. Nishio, and M. Morikura, "Reinforcement learning based predictive handover for pedestrian-aware mmwave networks," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 692–697, IEEE, 2018.

[35] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. 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.

[36] 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.

[37] G. Brown, "Exploring the potential of mmwave for 5g mobile access," tech. rep., Qualcomm, 2016. available at https://eu-ems.com/event-images/Downloads/Qualcomm20Whitepaper.pdf.

[38] W. Roh, "5g: from vision to reality," in *The Silicon Valley 5G Summit*, Samsung Electronics, 2018. available at https://www.researchgate.net/publication/243771074_DIY_Corpora_the_WWW_and_the_Translator.

[39] J. Bains, "Platform based design accelerates 5g test," in *The Silicon Valley 5G Summit*, National Instruments, 2016. available at http://www.samsung.com/global/business/networks/events/Silicon-Valley-5G-Summit/attachments/S4-NI-Jin-Bains.pdf.

[40] G. R. MacCartney, H. Yan, S. Sun, and T. S. Rappaport, "A flexible wideband millimeter-wave channel sounder with local area and nlos to los transition measurements," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–7, IEEE, 2017.

[41] S. Sur, X. Zhang, P. Ramanathan, and R. Chandra, "Beamspy: enabling robust 60 ghz links under blockage," in *13th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 16)*, pp. 193–206, 2016.

[42] T. Wei, A. Zhou, and X. Zhang, "Facilitating robust 60 ghz network deployment by sensing ambient reflectors," in *14th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 17)*, pp. 213–226, 2017.

[43] A. Zhou, X. Zhang, and H. Ma, "Beam-forecast: Facilitating mobile 60 ghz networks via model-driven beam steering," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pp. 1–9, IEEE, 2017.

[44] 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.

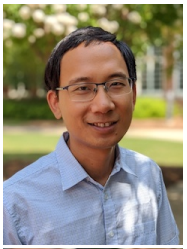
[45] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2-3, pp. 235–256, 2002.

[46] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in *Proceedings of the 19th international conference on World wide web*, pp. 661–670, ACM, 2010.

[47] D. Bouneffouf and I. Rish, "A survey on practical applications of multi-armed and contextual bandits," *arXiv preprint arXiv:1904.10040*, 2019.

[48] M. L. Attiah, M. Isa, A. Awang, Z. Zakaria, N. F. Abdullah, M. Ismail, and R. Nordin, "Adaptive multi-state millimeter wave cell selection scheme for 5g communications," *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 8, 2018.

[49] H. Tabassum, M. Salehi, and E. Hossain, "Fundamentals of mobility-aware performance characterization of cellular networks: A tutorial," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2288–2308, 2019.

**Li Sun** is currently a Ph.D. student in the Department of Computer Science and Software Engineering at Auburn University. He received the M.S. and Ph.D. degrees in Systems Engineering from Southeast University, China in 2008 and 2013, respectively. He used to be a visiting scholar in RWTH Aachen University, Germany in 2010. His research interest includes optimization in wireless network and machine learning.

**Jing Hou** is currently a Ph.D. student in the Department of Computer Science and Software Engineering at Auburn University. She received the B.S. degree in Computing Science from Nanjing Tech University, Nanjing, China in 2004 and the Ph.D. degree in Systems Engineering from Southeast University, Nanjing, China in 2011, respectively. Her research interest includes game theory and network economics.

**Tao Shu** is currently an associate professor in the Department of Computer Science and Software Engineering at Auburn University. He received his Ph.D. in Electrical and Computer Engineering from The University of Arizona in 2010. He received the B.S. and M.S. degrees in Electronic Engineering from the South China University of Technology, Guangzhou, China in 1996 and 1999, respectively, and the Ph.D. degree in Communication and Information Systems from Tsinghua University, Beijing, China in 2003. Prior to his academic position, he was a senior engineer in Qualcomm Atheros Inc. from Dec. 2010 to Aug. 2011. His research aims at addressing security and performance issues in wireless networking systems, with strong emphasis on system architecture, protocol design, and performance modeling and optimization.