IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 22, 2020, accepted July 16, 2020, date of publication July 21, 2020, date of current version July 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3011015

Neighbor Cell List Optimization in Handover Management Using Cascading Bandits Algorithm

CHAO WANG[1], JIAN YANG[1], (Senior Member, IEEE), HUASEN HE[1], RUIDA ZHOU[2], (Graduate Student Member, IEEE), SHUANGWU CHEN[1], AND XIAOFENG JIANG[1], (Member, IEEE)

[1]Department of Automation, University of Science and Technology of China, Hefei 230026, China
[2]Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA

Corresponding author: Huasen He (hehuasen@ustc.edu.cn)

This work was supported in part by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant ZDRW-KT-2016-02, in part by the Youth Innovation Promotion Association of the Chinese Academy of Sciences under Grant CX2100107001, in part by the Anhui Provincial Natural Science Foundation under Grant 1908085QF266, and in part by the China Electronics Technology Group Corporation (CETC) Joint Advanced Research Foundation under Grant 6141B08080101.

ABSTRACT Frequent handover is a key challenge in 5G Ultra-Dense Networks (UDN). In this paper, we show the significance of configuring Neighbor Cell List (NCL) in handover procedure. To cope with the high dynamic of UDN, we propose an online-learning method, namely the Cost-aware Cascading Bandits NCL configuration (CCB-NCL) algorithm, which applies the cascading model and Multi-Armed Bandits (MAB) theory to configure the efficient Neighbor Cell List (eNCL) and improves the handover performance by assisting the User Equipment (UE) to choose the optimal target Base Station (BS). We provide rigorous proof of regret bound to show the asymptotic convergence of the proposed CCB-NCL algorithm. The robustness and efficiency of the proposed algorithm are both demonstrated in different network scenarios, where varies BS densities, BS dynamic and network heterogeneity are considered respectively. In the simulation work, we reproduce two existing methods of configuring NCL in handover management, named dynamic threshold based solution and received signal strength based solution. In comparison with the existing solutions, the proposed algorithm can reduce the overlarge signaling cost and unnecessary delay in the preparation phase of handover procedure by significantly shortening the length of NCLs and reducing the number of scanned BSs. Extensive simulations are conducted in different scenarios to validate the robustness of the proposed algorithm and the results show that the proposed CCB-NCL algorithm is a superior approach to efficient handover management.

INDEX TERMS Cascading bandits, handover management, NCL configuration, user association, ultra-dense networks.

I. INTRODUCTION

To meet the growing demand of high data rate, low latency and high energy-efficiency in wireless networks, many new technologies have been proposed and applied in the fifth generation wireless communication (5G) [1]. Most of them focused on the 5G New Radio (NR) aspect, as well as the Next Generation Core (NGC) architecture [2]. To provide higher data rate with lower cost in wireless networks, the deployment of Small Base Stations (SBS) becomes one of the most promising technologies. For User Equipments (UE), higher data rates are always desired in modern wireless applications, such as Virtual Reality (VR), Augmented Reality (AR), real-time and streamed multimedia, etc. [3], [4]. Therefore, the small cell related technologies become more wildly applied recently, due to the distinct features of SBS networks.

Moreover, the advantages of SBS promote high density deployment of femtocells in Ultra-Dense Networks (UDN). Essentially, the deployment and maintenance cost of SBSs are much lower than that in macro cellular networks, while high data rate can be provided by SBS [3]. From the perspective

The associate editor coordinating the review of this manuscript and approving it for publication was Oussama Habachi.

of the wireless operator, the deployment of a large number of SBS will not only offload huge amounts of traffic from macro BSs to SBSs, but also will reduce maintenance and operation costs, and enhance the reliability of cellular networks [5]. However, the large scale and dense deployment of SBSs in wireless network brings several challenges, while handover management is one of the most challenging issues.

Handover management promotes the continuity of the connection between Base Stations (BS) and UEs during the movement of UEs. In mobile-assisted handover protocols, the pilot channel quality of neighboring BSs is measured by mobile UEs and reported to the serving BS. In this setting, handover may be triggered due to the deterioration of signal quality of current serving BS, or the events that one of the neighboring BSs can provide better signal quality. To improve the handover performance, many works on handover management have been conducted from the perspectives of industry and academic fields. The Next Generation Mobile Network (NGMN) association and the 3rd Generation Partnership Project (3GPP) have contributed a number of standards in wireless networks. Specifically, the scheme named Automatic Neighbor Relation (ANR) has been standardized in [6] and the handover procedure in NR system has been introduced in [2]. These schemes were proposed to improve operation and maintenance in wireless network, as well as improve the handover performance. Based on ANR and NCL, the handover target BS is determined by the BSs designing policy and management policy of operators. The target BS should be restricted to a specific neighbor BS with good channel quality so as to prevent handover failure. The NCL provides a specific set of neighboring BSs. In this case, if the target neighbor BS is excluded from the NCL, the handover performance will be degraded due to missing neighbors. Moreover, [7] showed that missing target BS in NCL has significant impacts on the mobility robustness and load balancing in mobility management.

To improve handover performance in UDN, NCL configuration is regarded as a significant task [8]. However, the dense deployment of BSs in 5G networks results in that NCLs include redundant BSs, which increase the signaling cost and time consumption on scanning the candidate BSs. Intuitively, removing the unnecessary BSs in NCL will improve the handover performance by providing lower signal overhead, faster handover, and lower energy consumption [9]. Since the NCL is used to control the scale of cell measurements in handover management, the mobile UE should receive the information about the pilot channels from potential neighbor BSs in NCL, rather than all potential BSs [10]. The NCL-based solution consumes much less time to acquire necessary measurements. More specifically, in a NCL-based solution, each UE is presented with a NCL which only contains necessary candidate BSs. Then, the mobile UE searches for the target BS by monitoring only the pilot signal quality of the candidate BSs in NCL. From this perspective, the configuration of NCL plays an important role in handover management. A NCL should contain enough potential BSs to ensure that at least one BS can satisfy all the handover requirements. However, two facts make UE suffer from excessive latency before finding the target neighboring BS. (1) the ultra dense deployment of SBSs makes the scale of NCL over large; (2) the measurement capacity of mobile UE is limited. And then excessive latency will cause call drops, packets missing or other degradation on handover performance.

A. RELATED WORKS

In some related works, NCLs were configured manually by estimating the signal propagations at the beginning of the network deployment with toolkit [11], [12]. Intuitively, the configuration of NCL was simply selecting all cells overlapping with the serving BS according to the network topology. However, it is quite different in practice, where the radio coverage will be influenced by the environment. Consequently, obtaining accurate predictions of BS coverage is very complex. The problems in configuration of NCL includes dealing with missing neighbors, Physical Cell Identifier (PCI) confusion, ultra-distant neighbor cell, redundant neighbor cell, etc. [13]. In the manual configuration of NCL, static information, such as BS location, antenna pattern, and received signal pattern, was used to predict cell coverage and neighbor relations. For example, the authors in [13] assigned different weighted factors to multiple static parameters of the BSs, and manually configured the NCL with the descending neighboring coefficients. Therefore, the handover procedure was performed with high complexity [12], and the static predictions were always lack of accuracy and could not capture the sensitivity and dynamics of radio propagation conditions in real networks.

Some works on configuration of NCL have been conducted by using knowledge of real network parameters. Initially, a BS needs to sense the spectrum and discover all BSs in its neighborhood, and then the NCL for this BS is derived automatically. The key problem in the sensing phase is determining whether a neighbor cell should be added to NCL. Generally, the NCL should adapt to the changes of transmission characteristics (e.g., transmitting power, distance between BSs), and the cell radius also should be considered [8]. In some special cases, the neighbor BSs may be blocked by building or other occlusions, this problem has been solved in [5], [14]–[16]. To mitigate the impact of disappearing BS, the authors in [17] proposed a NCL optimization method by considering the fade duration outage probability. Considering both the statistic of previously visited cells and the estimation of distance between BSs, the authors in [16] proposed a principle of obstructed paths to help UE discover the neighboring BSs.

Once the NCL is initialized in UDN, it must be optimized due to its overlarge size. In [18] and [19], the length of NCL was reduced by detecting the BSs with low level of pilot channel and removing them from NCL. Taking the environmental changes into consideration in NCL optimization, authors in [10] presented a new method including a self-configuration phase and a self-optimization phase, where the measurements reported by UE were used to optimize the NCL automatically. In [20] and [21], the authors exploited the modification in the scanning interval of individual cells, and proposed a dynamic adaptation of NCL based on considering two criterions including the probability of handover to a specific cell and Received Signal Strength (RSS) observed by the UE. In [22], the measured RSS in the handover to femtocells was recorded by its serving macrocell.

B. MOTIVATIONS

For mobile UEs in cellular networks, handover management allows them to switch network access points, while guaranteeing the QoS and maintaining service continuity. However, the high dynamic and heterogeneity of 5G UDN make the handover management a critical challenge. In the handover procedure proposed by 3GPP for 5G networks, there are still several works to be done in some specific real-life scenarios where ultra-dense deployment of BSs is considered [23]. It is worth noting that the configuration of NCL plays an important role in handover management, which is used to ensure that the UE can be switched to the best target BS in an idle state. Therefore, the adjustment and optimization of NCLs are the key to guarantee the handover performance [13], [17].

NCL provides candidate BSs for handover, and the optimization of existing list is critical which has not been addressed thoroughly. Especially, the cell load information has not been fully utilized. For instance, in some related works [10], [18]–[21], the configuration of NCL was mainly based on RSS from neighboring BSs. This information was provided by the measurement reports from the corresponding UE. However, these handover management schemes considered only RSS for making the handover decision, without considering potential overloading of the target BSs, which can be optimized in configuration of NCL to enhance the handover performance [4]. Moreover, these works did not take the long-term performance knowledge into consideration, which means that the knowledge of previous handover decision is omitted in configuration of NCLs.

Learning algorithm can utilize the statistic information of previous handover results and provide an efficient solution to optimize NCL. Motivated by this, in this paper, we adopt the bandits algorithm in reinforcement learning theory to make full use of long-term knowledge of previous handover decisions, which can reflect the dynamic characteristics of radio propagation conditions. Moreover, in our proposed algorithm, the configuration of NCL is based on both RSS and the remaining capacity of the candidate BSs. Compared with the solutions proposed in previous works which only considered the instant RSS, the solution proposed in this work performs better in reducing the signaling cost and the latency caused by scanning overloading BSs.

C. CONTRIBUTIONS AND STRUCTURE

In our work, we consider both RSS and cell load of potential target BSs and use a learning method based on the long-term performance to configure the optimal NCL, which can avoid some unnecessary scans of BSs in NCL. More specifically, three main components are considered in the proposed approach of efficient handover management. First, we propose building an *efficient NCL* (eNCL) for each BS from a *ground-set NCL*. The eNCL provides UE with a list of potential handover targets sorted according to their qualities, and the size of the list is limited. The ground-set NCL is the list of all potential BSs and its size can be significantly larger than that of the eNCL. Second, we adopt a cascading model [24] for the configuration of eNCL. In this model, the UE is recommended with the eNCL from its serving BS. UE examines the list from the top to the bottom, and measures each candidate BS one-by-one. UE will settle on the first ‘‘attractive’’ BS, whose measurement, including both RSS and cell load, is above the handover threshold (possibly with some padded margin). Lastly, the configuration of eNCL is done *online*, via iterations with UE handovers. This naturally poses as a bandit problem, therefore we leverage the cost-aware cascading bandits [25] in the algorithm design. However, unlike the simple cascading bandits setting in [25], UDN based on small BSs (such as femto and pico), often encounter delay or missing UE measurements, based on which we have modeled the features of reward and cost on the proposed algorithm. The main contributions of this paper are summarized below.

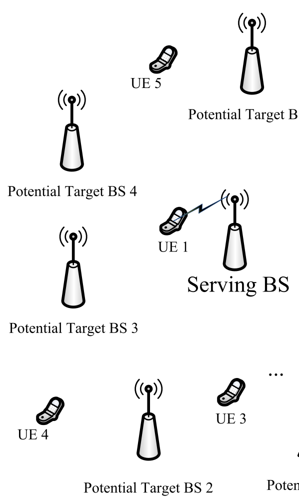
1) We propose the conception of eNCL, to handle the over large size of NCL in UDN. Based on this novel conception, we take the long-term performance knowledge into consideration, and present a cascading bandits framework for handover process.
2) We model the scanning latency as cost to the value function in the online learning process, and thus force the eNCL to reduce the latency in handover preparation phase by avoiding unnecessary BS scans.
3) We propose the Cost-aware Cascading Bandits NCL configuration (CCB-NCL) algorithm to dynamically and resourcefully configure the eNCL with convergence guarantee. Furthermore, finite-time analysis of the proposed algorithm suggests a sublinear regret behavior.
4) We conduct series of simulations of different network scenarios to evaluate the proposed handover scheme, and the results show the superiority of the proposed CCB-NCL algorithm compared to the existing methods.
5) Finally, we demonstrate the robustness and efficiency of the proposed algorithm in different network scenarios considering varies BS densities, BS dynamic, and network heterogeneity, respectively.

The rest of the paper is organized as follows. The system model is presented in Section II. In Section III, the handover management problem is formulated with cascading bandits theory. Section IV introduces the CCB-NCL algorithm, along with a convergence analysis of the proposed algorithm. The simulation results are presented in Section V. Finally, Section VI concludes the paper.

Through this paper, we use $\mathcal{L}$ to represent a NCL. $|\mathcal{L}|$ denotes the number of BSs in $\mathcal{L}$, and $\mathcal{L}(n)$ represents the n-th BS in $\mathcal{L}$. For the random variable P, its observation and expectation value are represented as $\hat{P}$ and $\bar{P}$, respectively. The notation $\mathbb{1}\{\cdot\}$ is used to denote the indicator function. Expectation operation is represented by $\mathbb{E}[\cdot]$. Finally, the notation $Pr\{\cdot\}$ is used to denote the probability of some events.

II. SYSTEM MODEL

In this paper, we consider an ultra dense cellular network, which contains a set of BSs (the average BS density is in the range of 128 to 512 BSs $/km^2$) and a random number of UEs distributed according to a homogeneous Poisson Point Process (PPP) with intensity λ_P. For heterogeneous networks, the set of BSs, $\mathcal{N} = \{1, \ldots, N\}$, consists of macro BSs and femto BSs with different coverage and capabilities. It is assumed that each type of BS has the same transmit power, ie. $\mathcal{P}_{macro}$ or $\mathcal{P}_{femto}$. And the association regions for the BSs can be assumed as a weighted voronoi tessellation [26]. Considering the random movements of UEs, we assume the number of UEs in the coverage region of a specific BS follows the Poisson distribution with mean $\bar{N} = \lambda_P S$, where S is the area of the coverage region.



NCL	Efficient NCL
BS 1	BS 3
BS 2	BS 4
BS 3	BS 2
BS 4	
BS 5	
BS 6	

FIGURE 1. Example of user association in UDN.

With the depicted system model, we give an example of user association in UDN as shown in Fig. 1, where UE 1 moves within the cell coverage area of its serving BS. UE 1 keeps monitoring the serving channels, and transmits the measurement reports to its serving BS. Considering core network functionalities, the cells are connected to the Evolved Packet Core (EPC), more specifically to the Access Mobility Function (AMF) and the Serving Gateway (S-GW). Furthermore, the BSs are interconnected via Xn interface, which enables the BSs to directly communicate with each other and perform functionalities such as handover [2]. Therefore, the users in the network can handover among different cells.

The handover type considered in this paper is described as the Xn based inter NG-RAN handover procedure in [2], which starts with measurement reports from the UE to its serving BS [27]. When the UE is triggered for handover, both RSS and the cell load are used as criteria to select the target BS. RSS observed by the UE is used to evaluate the channel quality of a candidate BS, and we assume RSS of the target BS should be higher than a fixed threshold (i.e. −70dBm in this work) to provide stable channels [28]. Besides, we define the cell load as the ratio of the number of connected UEs to the maximum number of UEs can be served, and the cell load should not exceed 3/4 [4].

As it was mentioned in [2], the handover procedure contains three phases: handover preparation, handover execution and handover completion. This paper focus on the first phase, where a NCL is generated for UE. Once the handover is complete, the performance is evaluated using some indicators such as signalling cost and handover latency caused by scanning candidate BSs, and the operation repeats in the next slot. The sequence of operations within each slot can be illustrated in Fig. 2.

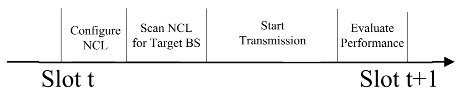


FIGURE 2. The handover operation for UE in slot t.

III. PROBLEM FORMULATION

A. CASCADING BANDITS MODEL

In this part, we adopt the cascading bandits model [24] to formulate the NCL optimization problem. In this model, we consider a K-armed stochastic bandit system, and all BSs are represented by the arms in ground set $\mathcal{K}$. Specifically, the learning agent interacts with the environment, where $|\mathcal{K}|$ BSs are distributed, and estimates the statistics of the BSs. When UE is triggered for a handover process at time t, the agent recommends an *eNCL*, denoted as $\mathcal{L}_t = \{\mathcal{L}_t(1), \mathcal{L}_t(2), \ldots, \mathcal{L}_t(|\mathcal{L}_t|)\} \in \Pi_{|\mathcal{L}|}(\mathcal{K})$, from the ground set $\mathcal{K}$. Note that $|\mathcal{L}|$ can be much smaller than $|\mathcal{K}|$. Then UE examines the recommended list $\mathcal{L}_t$ sequentially. UE chooses the first suitable candidate BS based on RSS and cell load. Denote $\tilde{\mathcal{L}}_t$ as the list of BSs that have been actually examined in slot t, $\tilde{\mathcal{L}}_t \subseteq \mathcal{L}_t$, and then we have that the target BS is the last BS in $\tilde{\mathcal{L}}_t$. If the handover requirements could not be satisfied by any BS in eNCL, the handover fails and the UE suffers call drops. Then, the above procedure will repeated in next time slot.

B. COST-AWARE CASCADING BANDITS

To improve the handover performance, it is very critical to avoid unnecessary BSs scanning. Thus, we need to reduce the length of NCL, and adjust the order of BSs in the presented eNCL. To achieve more accurate evaluation of BSs, we take both channel quality and cell capacity of BSs into consideration in NCL optimization. In order to explicitly enforce the constraint of delay in handover preparation phase, we have modified the cost-aware cascading bandits framework proposed in [25] and modelled the number of scanned BSs as a random cost.

To evaluate the presented eNCL at each time slot t, we adopt the conception of *reward* in bandit theory. In handover process, the straightforward reward that learning agent received is relevant to the distributions of the RSS and cell load of the target BS. Specifically, the *reward* depends on the sample value of RSS and instantaneous available capacity of the target BS at time slot t. Before finding the target BS, the number of scanned BSs in the presented NCL is formulated as the *cost* in this time slot. Combining the rewards and costs, we hence have the evaluation of the presented NCL at each time slot t. After several learning iterations, statistical values can be used to approximately describe the unknown RSS and cell load of BS, and the agent could finally present UE with an eNCL.

For the UE triggered for handover, the received power from the i-th BS is denoted as $P_{i,t}$, which is dependent on path loss and small-scale fading at time slot t,. In this work, we adopt Rayleigh fading as small-scale fading between UE and BS. At time slot t, $P_{i,t}$ is a sample value from a given distribution. Thus, the expected value of $P_{i,t}$ can be expressed as

$$\bar{P}_i = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} P_{i,t} \tag{1}$$

Similarly, we formulate the instantaneous cell load of the i-th BS in the arbitrary time slot t as a sample value $\eta_{i,t}$, which is determined by the number of UEs associated to the i-th BS. Due to the random distribution of mobile UE, we use a homogeneous PPP to model the locations of UEs. It is reasonable to assume that in the arbitrary time slot t the number of UEs served by the i-th BS follows a Poisson distribution with mean number equals to the average number of UEs located in the coverage area of the i-th BS. Then the linear correlation is adopted between the cell load and the number of UEs associated to the BS. Therefore, we have the expected cell load of i-th BS as

$$\bar{\eta}_i = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \eta_{i,t} \tag{2}$$

Definition 1: We use r_t to represent the straightforward reward received by the learning agent at time slot t. After scanning $|\tilde{\mathcal{L}}_t|$ BSs in the presented NCL, the target BS $\mathcal{L}(|\tilde{\mathcal{L}}_t|)$ is chosen at time slot t. Considering the RSS and cell load of the target BS, the straightforward reward is defined as

$$r_t = \frac{P_{|\tilde{\mathcal{L}}_t|,t}}{\eta_{|\tilde{\mathcal{L}}_t|,t}} \tag{3}$$

For a given NCL $\mathcal{L}_t$, we use $\tilde{\mathcal{L}}_t$ to denote the subset of BSs scanned until finding the target BS. Then, the length of $\tilde{\mathcal{L}}_t$, denoted as $|\tilde{\mathcal{L}}_t|$, is random and depends on the observed $P_{i,t}$, $\eta_{i,t}$ as well as the stopping condition (e.g., one scanned BS satisfies all the requirements of handover).

Definition 2: The cost in the evaluation of the NCL at time slot t is denoted as $cost_t$, and is set to be proportional to the number of scanned BSs in $\mathcal{L}_t$. It can be written as

$$cost_t = \beta|\tilde{\mathcal{L}}_t| \tag{4}$$

where β represents a constant coefficient.

Considering the above definitions of reward and cost, the *net-reward* received by the learning agent at time slot t is defined as

$$reward_t = \frac{P_{|\tilde{\mathcal{L}}_t|,t}}{\eta_{|\tilde{\mathcal{L}}_t|,t}} - \beta|\tilde{\mathcal{L}}_t| \tag{5}$$

and the *per-step regret* at time slot t can be defined as:

$$reg_t = reward^* - reward_t \tag{6}$$

where $reward^*$ denotes the optimal net-reward that could be obtained with the optimal policy at time slot t. If the statistics of $P_{i,t}$, $\eta_{i,t}$ are known beforehand, the optimal NCL $\mathcal{L}^*$ is determined, also the optimal reward can be generated according to the same stopping condition.

The optimal per-step reward $reward^*$ can be written as

$$reward^* = \frac{\bar{P}_{i^*}}{\bar{\eta}_{i^*}} - \beta \tag{7}$$

where the optimal NCL $\mathcal{L}^*$ includes BSs with $\frac{\bar{P}_i}{\bar{\eta}_i} > \tau$, and ranks the BSs in a descending order of $\frac{\bar{P}_i}{\bar{\eta}_i}$. The optimal target BS i^* (sometime may not be the handover target BS) is placed at the fist position in $\mathcal{L}^*$.

Without prior statistics of $\{P_{i,t}, \eta_{i,t}\}$, the learning process of these two distributions can be viewed as a bandit problem. Specifically, taking the observation in learning steps into consideration, we aim to configure $\mathcal{L}_t$ to minimize the *cumulative regret* over time horizon T, which is given by

$$R(T) = \sum_{t=1}^{T} \mathbb{E}[reg_t] = T \cdot reward^* - \sum_{t=1}^{T} \mathbb{E}[reward_t] \tag{8}$$

IV. NCL OPTIMIZATION ALGORITHM DESIGN

With the objective of overcoming the limitations of the RSS-based handover management, this paper aims at designing a more efficient NCL configuration algorithm in UDN, and improving the handover performance. The main idea is to provide UE with a sorted BSs list with limited length. As mentioned in Section III-B, the handover management problem formulated in a cascading bandits framework calls for a direct application of the proposed cascading bandits algorithm in [24]. However, such a straightforward application ignores some practical constraints in user association problem in UDN. In this section, we highlight several limitations in UDN, and discuss how to design the eNCL configuration algorithm to handle the NCL optimization problem.

A. CCB-NCL ALGORITHM

In this section, we address the user handover problem with bandit theory. More specifically, we are interested in finding a NCL configuration strategy, which can minimize signaling cost and handover delay. Based on the formulation of the cascading bandits model in Section III-B, we now have an exploration versus exploitation dilemma which is a trade off between exploring the environment to find advantageous actions, and taking the empirically optimal action based on current observations. Actually, this scenario can be viewed as searching for a balance between presenting currently optimal NCL with those most potential target BSs, and sampling on those BSs with less estimation. Successive presentation of NCL yields rewards, which are independent and identically distributed according to the unknown distribution of RSS and cell load. The most common ideas of bandit algorithms to estimate unknown distribution come from Upper Confidence Bound (UCB) policy. In the UCB policy, an upper confidence index is associated to each item in the ground set. This index represents a biassed evaluation of each item, and is generally complex to calculate. For a specific item, the upper confidence index depends on the entire sequence of rewards and costs generated in the learning process. In each step, the policy uses it as an estimation for the corresponding reward expectation, and picks the item with the current highest index for the next play [29].

One critical limitation of applying the algorithm proposed in [25] to handover management, is that handover delay could not be taken into consideration. We are interested in finding a NCL configuration algorithm that minimize signaling cost and handover delay. It is worth noting that by including the scanning cost $cost_t = \beta\tilde{\mathcal{L}}_t$ in the learning process, the proposed NCL configuration algorithm can balance the tradeoff between exploitation and exploration of potential target BS, while minimize the number of scanned BSs and the total length of presented NCL.

In order to address the NCL configuration problem in handover process, based on the UCB policy, we propose the Cost-aware Cascading Bandits NCL configuration (CCB-NCL) algorithm, which consists of a initialization phase and a iteration phase. During the initialization phase of the learning process, UE arrivals at a specific location on the edge of the coverage area of serving BS, and handover is triggered. Here, UE is presented with the NCL that contains all potential BSs. UE scans each BS in NCL sequentially, until finding a BS that satisfies all handover requirements. For the i-th scanned BS, UE measures the RSS P_i, and checks the cell load η_i. After finding the target BS, these measurements and other metrics such as the number of scanned BSs are transmitted to MME (Mobility Management Entity).

In the t-th iteration, MME evaluates all potential BSs based on the measurements recorded in the previous iteration. Using the UCB policy mentioned above, the upper confidence indexes of all BSs are calculated. The eNCL $\mathcal{L}_t$ includes BSs with upper confidence index no less than the threshold, and ranks these BSs in a descending order of upper confidence index. Then, UE scans the BSs sequentially until a handover target BS is found. MME receives the sample values of RSS and cell load of each scanned BS, and updates the estimations of these two distributions of BSs iteratively. The reward and cost corresponding to $\mathcal{L}_t$ are calculated according to (3) and (4) respectively. Subsequently, in the $(t+1)$-th iteration, the handover is triggered again when a UE moves to the same location, and the learning process will repeat. The eNCL obtained after T iterations can be regarded as the optimal configuration of NCL associated with the specific location. Then, any UE moved to this location could be presented with the trained eNCL without any calculations.

The CCB-NCL is given with the pseudo code in Algorithm 1. Here, the total times that the k-th BS has been chosen by UE right before time t is tracked and represented by the variable $N_{k,t}$. The sample values of P_k and η_k at time slot t are denoted by $\hat{P}_{k,t}$ and $\hat{\eta}_{k,t}$, respectively. In this algorithm, we adopt $u_{k,t} = \sqrt{\frac{\alpha \log t}{N_{k,t}}}$ as the UCB padding term on the state of the k-th BS at time slot t, and the UCB parameter α is a positive constant no less than 1.5 [29].

Algorithm 1 Cost-Aware Cascading Bandits NCL Configuration Algorithm

Input: number of iterations T, UCB parameter α, UCB threshold τ;
Output: eNCL $\mathcal{L}_T$
Initialization: Select all BSs in $\mathcal{K}$, and observe their RSS and cell load;
Iteration: **while** $t < T$ **do**
  for $k = 1, 2, \ldots, |\mathcal{K}|$ **do**
    if $\frac{\hat{P}_{k,t}+u_{k,t}}{\hat{\eta}_{k,t}+u_{k,t}} > \tau$ **then**
      $k \to \mathcal{L}_t$
    end
  end
  Rank BSs in $\mathcal{L}_t$ in the descending order of $\frac{\hat{P}_{k,t}+u_{k,t}}{\hat{\eta}_{k,t}+u_{k,t}} > \tau$
  for $i = 1 : \mathcal{L}_t$ **do**
    Pull BS $\mathcal{L}_t(i)$ and observe $P_{\mathcal{L}_t(i),t}, \eta_{\mathcal{L}_t(i),t}$;
    if *BS in* $\mathcal{L}_t$ *is chosen* **then**
      break
    end
  end
  Update $N_{i,t}, \hat{P}_{i,t}, \hat{\eta}_{i,t}$;
  Calculate the cost $\beta\tilde{\mathcal{L}}_t$ and reward $reward(t)$ at t.
  $t = t + 1$
end

B. ALGORITHM ANALYSIS

In this section, we analyze the performance of Algorithm 1 in terms of the finite-time regret as a function of time.

We present a regret bound of the optimal eNCL in the following theorem.

Theorem 1: Denote $\Delta_i^2 = (\bar{P}_i - \bar{\omega}_i)^2$. When Algorithm 1 runs with UCB parameter α, and in a given time horizon T, the regret is upper bounded by

$$R(T) \leq \sum_{i \in K \setminus \mathcal{L}^*} \frac{(16K\alpha)\log(T)}{\Delta_i^2} + O(1). \tag{9}$$

Proof: The complete proof is shown in Appendix . The proof contains four main steps. First, we show that the regret comes from the event that there exists a BS i whose sample average of RSS or cell load lies outside the corresponding confidence interval, and the event that the BSs in list $\mathcal{L}_t$ are not ranked in correct order. Second, the regret at time t is decomposed into three parts according to different events. Third, we calculate the regret bound based on the number of times that each suboptimal item is chosen in T steps. Finally, we sum up the regret of all suboptimal items. ■

From Theorem 1, we can see that the extra cost associated with scanned number have no effect on the order of the regret upper bound. Intuitively, as bandit learning results gradually converge to the optimal eNCL, the additional cost caused by the unnecessary scanning of BSs diminishes and has no impact on the scaling of regret with respect to T. This indicates that the cost-aware cascading bandits model can be well adopted in handover management.

V. SIMULATION AND PERFORMANCE ANALYSIS

A. SIMULATION ENVIRONMENT

In this section, we conduct several sets of simulations to illustrate the advantages of the proposed handover management solution. We apply the proposed CCB-NCL algorithm in different scenarios, and accomplish the performance comparisons between the proposed algorithm and the existing solutions. Specifically, the performance of traditional RSS-based solution is adopted as the benchmark. We consider the UDN where the BSs are densely deployed in urban environment. To account the irregularity of the real networks, we adopt a Voronoi model rather than the ideal grid-based hexagonal cells model. In simulations, BSs are randomly distributed and the coverage regions of the BSs form a Voronoi tessellations. Unless otherwise specified, all BSs are with fixed transmit power. To simplify the simulation, we initialize the network where each UE associates with one of these BSs based on the RSS.

In this work, we simulate a 500 meters by 500 meters urban area, where there are 64 femtocell BSs with a fixed transmit power, i.e. $\mathcal{P}_{femto} = 23\ dBm$. Besides, the UEs are distributed according to a PPP with intensity $\lambda = 1/250$ [20]. We assume that UEs in this area move towards random directions, leaving or entering the coverage of a specific BS. Without loss of generality, we consider a generic UE (also referred to as the UE of interest), who moves to the cell edge of its serving BS. The statistic path loss model and the Rayleigh fading are used to calculated the RSS at the UE of interest, and the cell load of each BS is assumed to be linear to the number of UEs in its coverage area.

As for the pathloss model, we adopt the configuration of HetNet scenario in 3GPP specifications [30], and conduct the system-level simulation with the 3GPP pathloss model in small cells networks, which is given by

$$PL(d)[dB] = 15.3 + 37.6 \times \log_{10}(d) + L_{ow}, d > d_0. \tag{10}$$

All the other parameters of the general urban explicitly model are provided in Table 1.

TABLE 1. Simulation parameters.

Parameters	Value
Thermal noise density	-174dBm/Hz
UE noise figure	5.5dB
Carrier frequency	2.1GHz
Bandwidth	20MHz
Penetration loss (L_{ow})	10dB
d_0	1m

In order to verify the adaptability of the proposed algorithm, we simulate networks with different BS densities. Similarly, the robustness of the CCB-NCL algorithm is verified through simulations in the scenarios where dynamic BSs are considered. When the dynamic presence of BS is considered, it is assumed that the potential target BS e might be turned off with off-probability p_e. The off-probability of each BS is generated randomly in each iteration and unknown to UE. Moreover, we also verify the robustness of our proposed algorithm in heterogeneous networks where macro BSs and femto BSs are coexist and with different transmit power, i.e. $\mathcal{P}_{macro} = 35$ dBm and $\mathcal{P}_{femto} = 23$ dBm, respectively [16]. The cell capacity of each macro BS is four times of that in one femto BS. Unless otherwise specified, in all scenarios, the learning process iterates 10^4 times and the location of UE is regenerated for each simulation.

B. COMPETITIVE ALGORITHMS

We compare the proposed algorithm with two other NCL configuration schemes: (1) self-optimizing algorithm with dynamic threshold based on the previous handover information [20], [21]; (2) RSS-based NCL configuration method in the traditional handover management [6].

The first compared algorithm configures NCL based on the dynamic threshold, and is denoted as "DT-based solution". In this solution, the probability of handover to each BS and the SINR observed by the UE were used to select the candidate BSs [20], [21]. Under this setting, BSs with higher probability to be selected as the handover target, will be scanned more frequent. In the simulations using the first compared algorithm, the probability of handover to the given cell is calculated by the statistic of 10^3 handover results.

The second method is traditional "RSS-based solution" [6]. In this method, the potential BSs are chosen based on their instantaneous performance without leveraging prior information. The RSS-based NCL is configured with sample value of

RSS, and it includes BSs whose measurement of RSS is above the handover threshold. Once the UE moves to the cell edge and the channel deteriorates below a threshold, the handover procedure is triggered and the RSS-based NCL is presented to the UE.

C. PERFORMANCE METRICS

In order to illustrate the enhancement on handover performance using the proposed algorithm, we adopted four metrics including the length of NCL presented to UE, the number of scanned BSs before the handover target is chosen, the proportion of wrong cells in NCL, and the probability of choosing the optimal BS as handover target. The length of NCL is in proportion to the signaling cost in handover process, and the number of scanned BSs influences the handover delay in preparation phase significantly. Moreover, a cell in NCL is defined as a wrong cell if its RSS is lower than −70 dBm or the cell load exceeds 3/4, and the optimal BS is defined as the candidate BS with highest RSS and its cell load is less than 3/4 at the same time. Then the proportion of wrong cells and the probability of choosing the optimal BS as the handover target can reveal the performance of NCL configuration algorithm. We select these four metrics to illustrate the advantage of the proposed CCB-NCL algorithm on reducing signaling cost and handover delay.

D. SIMULATION RESULTS

In this section, we present the results of five experiments to reveal the superiority of the proposed algorithm in handover management, and compare its performance with both DT-based and RSS-based NCL configuration methods. In the first experiment, we have studied the impacts of CCB-NCL algorithm on the handover performance, in terms of the length of NCLs, and the number of scanned BSs. In the second experiment, we have analyzed the adaptability of the proposed algorithm considering different BS densities. The robustness of proposed algorithm is validated in the third experiments, where the dynamic property of BSs in a real networks is considered. Then, in order to evaluate the efficiency of the proposed algorithm, we have compared the performance of NCLs presented after different numbers of iterations. Lastly, we apply the proposed algorithm in a heterogeneous network, and compare its performance with the competitive solutions.

We first evaluate the performance of the proposed CCB-NCL algorithm and compare it with two competitive solutions. In Fig. 3, we plot the Cumulative Distribution Function (CDF) of the length of NCLs in 500 independent trials, using the proposed algorithm, DT-based and RSS-based methods, respectively. As it can be observed from Fig. 3, the length of NCL presented by the proposed algorithm is smaller than that in competitive solutions. By calculating the average length of NCLs obtained from three solutions, it can be found that the proposed algorithm has reduced the length of NCL by nearly 50% and 80%, respectively. This is caused by several facts. The density deployment of femtocell BSs results in large number of potential BSs that satisfy the RSS requirement. However, most of them are not necessary to be included in presented NCL, due to their limited physical capacities. Moreover, the target BS of handover is selected from several top-ranked BSs in the list, and there is no need to examine all the potential BSs on the remaining NCL. Thus, by excluding those potential BSs with high cell load, the proposed algorithm reduces the length of NCL significantly. Since the signaling cost is approximately proportional to the scale of the NCL, the proposed CCB-NCL algorithm outperforms both the DT-based and the RSS-based method in reducing the signaling cost during the handover preparation phase.

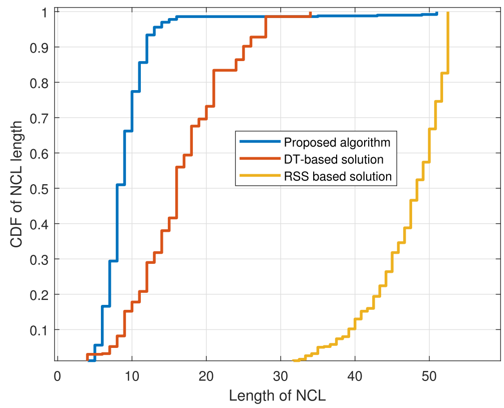


FIGURE 3. Comparison of NCL length between the proposed algorithm and competitive solutions.

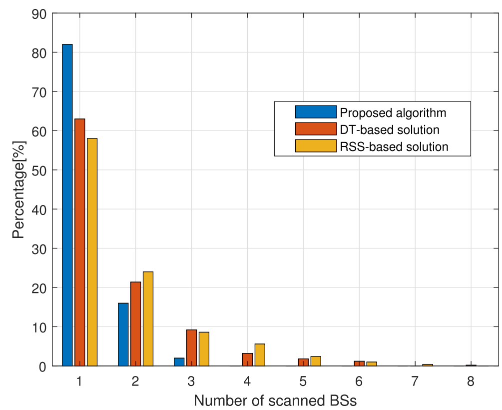
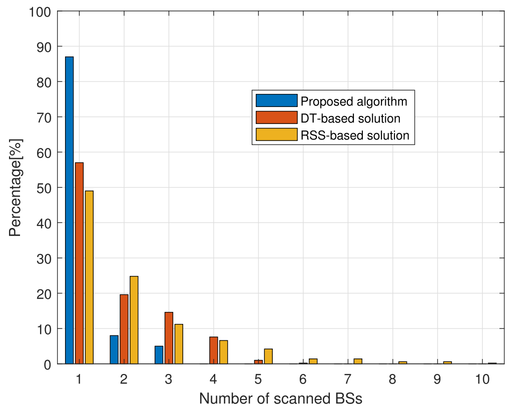


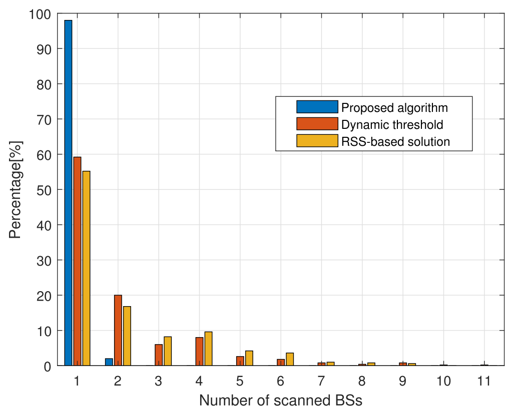
FIGURE 4. Comparison of the number of scanned BSs in NCLs obtained from different solutions.

Scanning of unnecessary BSs in the NCL incurs extra delay in the preparation phase of handover procedure. Thus, the distribution of the number of scanned BSs is exhibited in Fig. 4. It shows that the proposed CCB-NCL algorithm outperforms competitive solutions on the number of scanned BSs. Specifically, more than 80% of UEs can find the target BS after scanning the first BS in the eNCL, and most of them require to scan no more than 3 BSs. However, given a NCL generated by DT-based or RSS-based solution, only nearly 63% or 58%

UEs can achieve that performance. This is due to the fact that the proposed learning-based algorithm considers not only the channel quality of each BS, but also the cell load, which leads to a more accurate evaluation of the candidate BSs. Moreover, the eNCL presents BSs with higher potential to UE. Intuitively, the excessively redundant elements in other solutions consume unnecessary time and increase the length of NCL. Thus, the scanning time for the UE to identify the target BS in the preparation phase will increase. For the traditional solution based on RSS, the *instantaneous* channel qualities are used to rank the candidate BSs, which could not reflect the long-term performance of BSs. Considering the numbers of scanned BSs, the proposed CCB-NCL algorithm enhances the efficiency of finding the target BS, and outperforms the competitive solutions in reducing the delay in handover preparation phase.



(a) High BS density scenario.



(b) Low BS density scenario.

FIGURE 5. Comparison of the number of scanned BSs in NCLs obtained with different BS densities: (a) number of BSs = 128; (b) number of BSs = 32.

Secondly, we study the adaptability of the proposed algorithm by considering different BS densities. The numbers of BSs are set from 32 to 128 in different network sceneries, and the intensities of PPP used to describe the corresponding distribution of UEs are set from $\lambda_1 = 1/500$ to $\lambda_2 = 1/125$ [20]. It can be observed in Fig. 5, the average numbers of scanned BSs for three solutions decrease with BS density. More specifically, in the network with high BS density, the average numbers of scanned BSs has decreased from 3.2 to 1.5 using the proposed algorithm, while it decreases from 2.3 to 1.1 in low BS density setting. Thus, regardless of the changes of BS densities, the adaptability of the proposed algorithm is guaranteed, and the performance improvement of the proposed algorithm is more significant when the BS density increases.

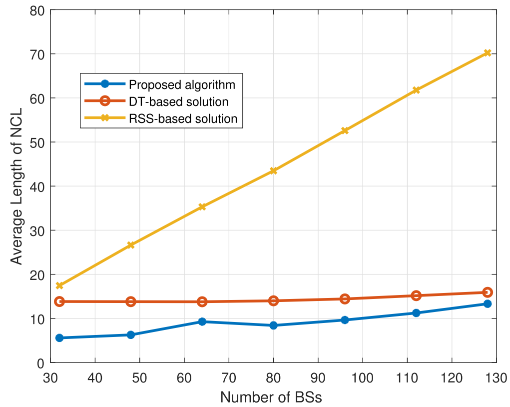


FIGURE 6. Impact of BS density on the average length of NCLs obtained from different methods.

Fig. 6 shows the impact of BS density on the average length of NCLs obtained from different methods. As it can be seen, the average length of NCLs presented by three solutions increases with BS density. Especially for the RSS-based solution, the length of NCL increases linearly and is much larger than those in other two solutions. This is due to the fact that there are more redundant BSs in ground-set NCL when the BS density increases. Even though the DT-based solution reduces the scale of NCL and keeps it in a lower level, our proposed algorithm provides more significant improvement. Compared with the RSS-based solution, the proposed algorithm reduced more than 60% of NCL length for different BS densities. Also, approximately 20% of the average length of NCL presented by DT-based solution can be reduced by the proposed algorithm. This result shows that the proposed algorithm is adaptive to the change of BS densities while reducing the scale of NCL significantly.

Besides, we also show the impact of BS density on the wrong cell proportion in NCLs in Fig. 7. It can be observed from Fig.7 that, the proposed algorithm outperforms other two solutions on the wrong cell proportion in all network scenarios. Specially, compared with the RSS-based solution, the proposed algorithm optimizes NCLs through reducing the wrong cell proportion by approximately 90%. For the DT-based solution, even though the wrong cell proportion decreases with the increasing of BS density, it still exceed that in the proposed algorithm for all network scenarios. This result shows that the proposed algorithm can optimize the NCL by excluding the BSs which do not satisfy handover requirements.

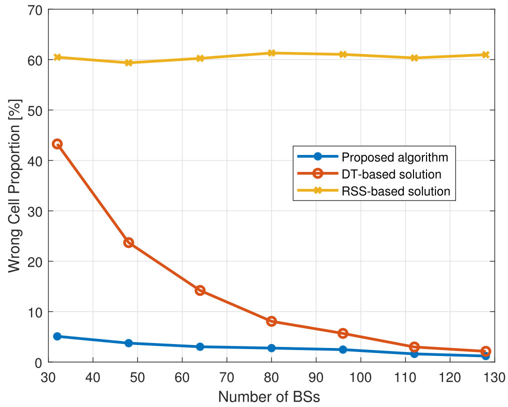


FIGURE 7. Comparison of the wrong cell proportion in NCLs for different solutions.

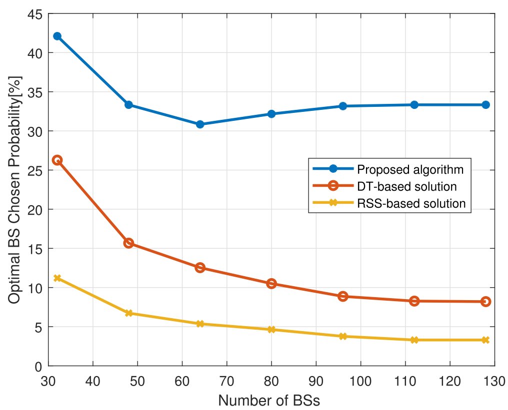


FIGURE 8. Comparison of the probability of choosing the optimal BS as the handover target for different solutions.

Likewise, as it can be seen from Fig. 8, the proposed algorithm outperforms other two solutions on the probability of choosing the optimal BS as the handover target in different network sceneries. Specially, compared with the DT-based solution and the RSS-based solution, the proposed algorithm improves the handover performance through increasing the probability by approximately 3 times and 7 times, respectively. In spite of that the optimal BS can provide highest RSS and sufficient cell capability, there still exist multiple sub-optimal BSs that also meet the handover requirements. This can explain why these three solutions keep the probability in the relatively low level. Moreover, with the increasing of BS density, the performance of both the DT-based and the RSS-based solution degrades due to the complexity of networks. However, the proposed algorithm maintains high probability of choosing the optimal BS as the handover target, even in ultra-high density networks, which shows that the proposed algorithm is adaptive to the changes of BS densities.

Thirdly, we validated the robustness of the proposed algorithm. Fig. 9 shows the performance of the proposed CCB-NCL algorithm in the scenario where the BSs will be turned off with the off-probability p_e. For example, the self-organization function triggered, power off, or other damage occurred on the BSs. Here, it is assumed that p_e is an independent and identically distributed (i.i.d.), and is

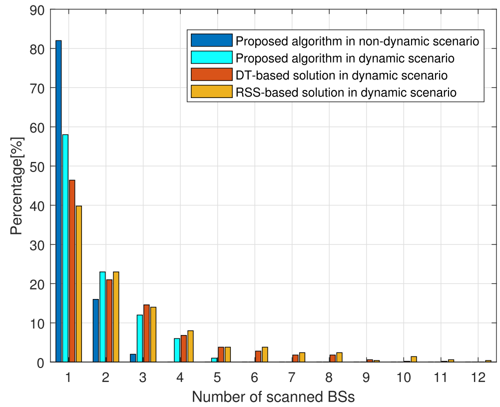


FIGURE 9. Impact of the dynamic BS on scanned number of BSs.

uniformly distributed within $[0, 0.3]$ [28]. This scenario is more practical as some BSs in the presented eNCL will no longer be detected when the scanning process begins. As it can be seen from Fig. 9, the performance deteriorated due to the disappearance of BSs in the presented eNCLs obtained from the proposed algorithm. However, the proposed algorithm still works well in keeping the number of scanned BSs in a low level.

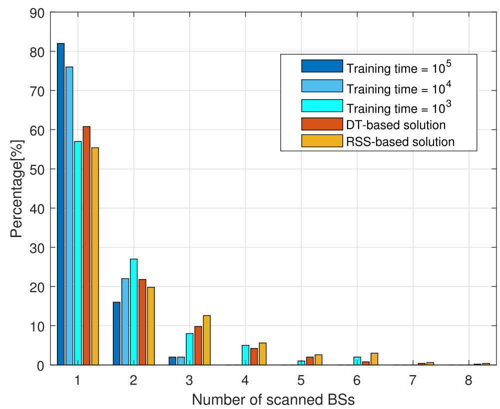


FIGURE 10. Impact of training times of the proposed learning algorithm on the number of scanned BSs.

Next, we evaluate the efficiency of the proposed algorithm with different number of learning iterations. The learning iterations are set to 10^3, 10^4 and 10^5 respectively. As it can be seen in Fig. 10, the average number of scanned BSs decreases with the number of learning iterations. However, even the learning algorithm is executed 10^4 iterations, the result still show the superiority of the proposed algorithm. It is worth noting that, comparing with results obtained from DT-based solution, the disadvantage of learning algorithm with 10^3 iterations appears. The intuitive explanation on this result is that the learning accuracy of the distributions of RSS and cell load depends on the iteration times of the proposed learning algorithm. Therefore, the long-term performance estimations on the potential BSs are quite important in the configuration of NCLs. After enough learning iterations, we have the eNCL associated with the location of UE. By recording all the presented eNCLs in the coverage area of one BS, we have a table that contains all UE positions associated with eNCLs. Since the distribution of channel quality and cell load would not change intensely, the table can be used to recommend NCL to UE without any extra learning cost over a period of time.

TABLE 2. Simulation results in heterogeneous network.

Solutions	Average NCL Length
RSS-based Solution	36.55
DT-based Solution	12.18
Proposed Algorithm	8.95

Finally, we apply the proposed algorithm to a heterogeneous network, where 4 macro BSs and 64 femto BSs are deployed. In this scenario, macro BSs are randomly distributed with a minimum distance $r_{min} = 300$ meters between each other, while the distance from any femto BS to one macro BS is set to be no less than 50 meters. As it can be seen from Table 2, the proposed algorithm provides significant improvement on reducing NCL length compared to the DT-based and RSS-based solutions. Moreover, the superiority of the CCB-NCL algorithm can be observed in Fig. 11. It can be concluded that our proposed CCB-NCL algorithm is robust in heterogeneous networks.

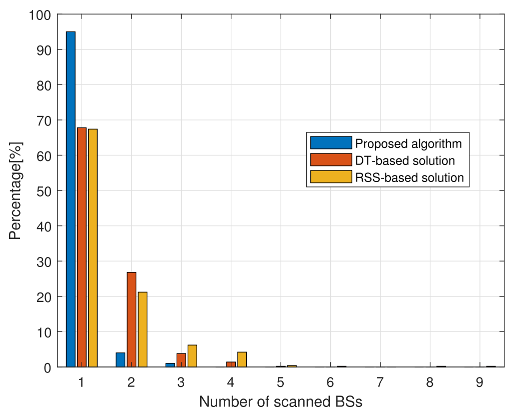


FIGURE 11. Comparison on the number of scanned BSs in NCLs obtained from different solutions in heterogeneous network.

VI. CONCLUSION

In this paper, we have studied the handover management in UDNs from the perspective of NCL configuration. By applying online learning methods, we use the long-term performance estimation instead of the instantaneous performance of BSs, to make the handover decision. We have proposed the CCB-NCL algorithms that optimizes the NCLs presented to UE when handover occurs. The proposed algorithm adopted the cascading bandits framework, and was proved with sub-linear regret bounds. We have made enhancements to the standard cascading bandits framework, so that the proposed CCB-NCL algorithm could enhance the handover performance. Extensive simulations have been conducted to show that the proposed algorithm can significantly shorten the length of NCLs and the number of scanned BSs in handover preparation phase, and reduce the wrong cell proportion while increasing the probability of choosing the optimal BS as the handover target in configuring NCLs, when compared with the existing solutions. Thus, we shown that the proposed algorithm can reduce the overlarge signaling cost and unnecessary delay in the handover preparation phase. Moreover, the robustness of the proposed algorithms was demonstrated in network scenarios with varies BS densities, BSs dynamic and heterogeneous BSs.

.

PROOF OF THEOREM 1

Let $reg(t)$ be the regret of the learning algorithm at time slot t. The event that the sample value of RSS $\hat{P}_i(t)$ or cell load $\hat{\eta}_i(t)$ of the i-th BS lies outside the corresponding confidence interval $[-u_{i,t}, u_{i,t}]$ at time t, can be denoted as $\xi_t = \{\exists i \in E, |\hat{P}_i(t) - \bar{P}_i| > u_{i,t}$ or $|\hat{\eta}_i(t) - \bar{\eta}_i| > u_{i,t}\}$; and $\bar{\xi}_t$ represents the complement of ξ_t. Let $\mathcal{B}_t = \{\exists i, j \in \mathcal{L},\ i < j,\ \frac{\bar{P}_i}{\bar{\eta}_i} < \frac{\bar{P}_j}{\bar{\eta}_j}\}$ be the event that the BSs in list $\mathcal{L}$ are not ranked in correct order. Run Algorithm 1, we can decompose the regret as

$$R(T) = \sum_{t=1}^{T} \mathbb{E}\left[[\mathbb{1}(\xi_t) + \mathbb{1}(\bar{\xi}_t)]reg(t)\right] \quad (11)$$

$$= \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{1}(\xi_t)reg(t)\right] + \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{1}(\bar{\xi}_t)\mathbb{1}(\mathcal{B}_t)reg(t)\right] + \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{1}(\bar{\xi}_t)\mathbb{1}(\bar{\mathcal{B}}_t)reg(t)\right], \quad (12)$$

since there is no regret for $\mathbb{1}(\xi_t)\mathbb{1}(\mathcal{B}_t)reg(t)$ and $\mathbb{1}(\xi_t)\mathbb{1}(\bar{\mathcal{B}}_t)$ $reg(t)$. Because once the parameters are estimated correctly, the ranking order of BSs in list is determined. Then, we need to bound these three terms in (12) respectively. For simplicity, we use the capital letter L instead of $|\mathcal{L}|$ to represent the length of the list $\mathcal{L}$ in this section.

The first term in (12) is relatively small because all of our confidence intervals hold with high probability. There are some BSs whose sample values of reward or cost lie outside the corresponding confidence interval. The regret resulted from these BSs is defined as

$$\sum_{t=1}^{T} \mathbb{E}[\mathbb{1}(\xi_t)reg(t)] \leq K + \sum_{t=K+1}^{T} \sum_{k\in\mathcal{K}} \Big(\Pr\left[|\hat{P}_k(t) - \bar{P}_k| > u_{k,t}\right] + \Pr\left[|\hat{\eta}_k(t) - \bar{\eta}_k|\right] \Big). \quad (13)$$

Then, we decompose the definition and rewrite (13) as

$$K + \sum_{k\in\mathcal{K}} \sum_{t=K+1}^{T} \sum_{n=0}^{t} \Big(\Pr\left[|\hat{P}_k(t) - \bar{P}_k| > \sqrt{\frac{\alpha \log t}{N_k(t)}}, N_k(t) = n\right] + \Pr\left[|\hat{\eta}_k(t) - \bar{\eta}_k| > \sqrt{\frac{\alpha \log t}{N_k(t)}}, N_k(t) = n\right] \Big) \leq K$$

$$+\sum_{k\in\mathcal{K}}\sum_{t=K+1}^{T}\sum_{n=1}^{t}2\exp\left(-2\frac{\alpha\log t}{n}\right)=K \tag{14}$$

$$+2\sum_{k\in\mathcal{K}}\sum_{t=K+1}^{T}t^{-2\alpha+1}\leq K+K\frac{\pi^2}{3}=\psi \tag{15}$$

where the (14) follows the Hoeffding's inequality.

To obtain the regret upper bound of the second part of (12), we first provide the following two lemmas.

Lemma 1: Let $p_s(e) = Pr\{P(t) > \tau_1,\ and\ \eta(t) > \tau_2\}$ be the probability that the BS e will be chosen at time t, where τ_1 and τ_2 are thresholds that the sample values of target BS should meet. For any BS $i \in \mathcal{K}$, if it is included in the NCL (i.e., $i \in \mathcal{L}_t$), then, it will be actually scanned in time frame t (i.e., $i \in \tilde{\mathcal{L}}_t$) with probability no less than

$$p_i = \frac{\prod_{j=1}^{K}(1-p_s(j))}{(1-p_s(i))}, \tag{16}$$

i.e.,

$$\Pr(i\in\tilde{\mathcal{L}}_t)\geq p_i\Pr(i\in\mathcal{L}_t). \tag{17}$$

Proof: According to Algorithm 1, for a BS included in $\mathcal{L}_t$, it will be scanned only when all of the former BSs were scanned and none of them was chosen as the target. Thus, the lowest probability that i-th BS will be chosen occurs in the worst scenario where the target BS lies on last position of $\mathcal{L}_t$. The corresponding probability is given by $\prod_{j=1}^{L-1}(1-p_s(j))$, which can be further bounded by p_i. ■

Lemma 2: We use $N_i(T)$ to represent the total number of times that the i-th BS has been selected as the target BS within T time slots. Then, for all $i \in \mathcal{L}$, we have

$$\mathbb{E}[N_i(T)]\geq p_i(T-\psi). \tag{18}$$

Proof: Based on the definition of $N_i(T)$, we have

$$\begin{aligned}\mathbb{E}[N_i(T)]&=\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(i\in\tilde{\mathcal{L}}_t)\right]\\&=\mathbb{E}\left[\sum_{t=1}^{T}(\mathbb{1}(\bar{\xi}_t)+\mathbb{1}(\xi_t))\mathbb{1}(i\in\tilde{\mathcal{L}}_t)\right]\\&\geq\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(\bar{\xi}_t)p_i\mathbb{1}(i\in\mathcal{L}_t)\right] \qquad (19)\\&=p_i\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(\bar{\xi}_t)\right]=p_i\mathbb{E}\left[T-\sum_{t=1}^{T}\mathbb{1}(\xi_t)\right]\\&\geq p_i(T-\psi) \qquad (20)\end{aligned}$$

where (19) follows Lemma 1, and (20) follows the upper bound on $\sum_{t=1}^{T}\mathbb{1}(\xi_t)$ in (15). ■

Then, the second part of (12) represents the case that the sample average of reward or cost of all BSs lie inside the corresponding confidence interval, but the BSs in the presented list $\mathcal{L}$ are not ranked in correct order. Then, the regret of this part is bounded as

$$\mathbb{E}\Big[\sum_{t=1}^{T}\mathbb{1}(\bar{\xi}_t)\mathbb{1}(\mathcal{B}_t)reg(t)\Big] \tag{21}$$

$$\begin{aligned}&\leq\sum_{j=2}^{L}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(\bar{\xi}_t)\mathbb{1}(\frac{\bar{P}_i}{\bar{\eta}_i}<\frac{\bar{P}_j}{\bar{\eta}_j})\right]\\&=\sum_{j=2}^{L}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(\bar{\xi}_t)\mathbb{1}\left(N_j(t)<\frac{4(1-\frac{\bar{P}_j}{\bar{\eta}_j})^2\alpha\log t}{(\frac{\bar{P}_{j-1}}{\bar{\eta}_{j-1}}-\frac{\bar{P}_j}{\bar{\eta}_j})^2\eta_j^2}\right)\right]\\&\leq\sum_{j=2}^{L}\mathbb{E}\left[\sum_{n=1}^{\Gamma_T}\mathbb{1}\left(N_j(\tau_n)<\frac{\alpha\log\tau_n}{\Omega_{j-1,j}^2}\right)\right] \qquad (22)\\&=\sum_{j=2}^{L}\mathbb{E}\left[\sum_{n=1}^{\Gamma_T}\mathbb{1}\left(N_j(\tau_n)<\frac{\alpha\log\tau_n}{\Omega_{j-1,j}^2}\right)\mathbb{1}\left(\tau_n\leq 2n\right)\right] \qquad (23)\\&\quad+\sum_{j=2}^{L}\mathbb{E}\left[\sum_{n=1}^{\Gamma_T}\mathbb{1}\left(N_j(\tau_n)<\frac{\alpha\log\tau_n}{\Omega_{j-1,j}^2}\right)\mathbb{1}\left(\tau_n>2n\right)\right] \qquad (24)\\&\leq\sum_{j=1}^{L}\left(\zeta_j+\frac{1}{2p_j^2}\right)+\rho=O(1) \qquad (25)\end{aligned}$$

where $\Omega_{i,j} = \frac{(\frac{\bar{P}_i}{\bar{\eta}_i}-\frac{\bar{P}_j}{\bar{\eta}_j})\eta_j}{2(1-\frac{\bar{P}_j}{\bar{\eta}_j})}$. Then, the second part of (12) is divided into (23) and (24), the former is bounded by

$$\begin{aligned}&\sum_{j=2}^{L}\mathbb{E}\left[\sum_{n=1}^{\Gamma_T}\mathbb{1}\left(N_j(\tau_n)<\frac{\alpha\log\tau_n}{\Omega_{j-1,j}^2}\right)\mathbb{1}\left(\tau_n\leq 2n\right)\right]\\&\quad\leq\sum_{j=2}^{L}\mathbb{E}\left[\sum_{n=1}^{\Gamma_T}\mathbb{1}\left(N_j(\tau_n)<\frac{\alpha(\log n+\log 2)}{\Omega_{j-1,j}^2}\right).\right] \qquad (26)\end{aligned}$$

Here, we define an independent random variable $Z_i(t)$, which can be written as

$$Z_i(t)=\begin{cases}0 & ,\ if\ \ \mathbb{1}(\bar{\xi}_t)=0\\ Ber(p_i) & ,\ if\ \ \mathbb{1}(\bar{\xi}_t)=1\end{cases} \tag{27}$$

With $Z_i(t)$, (26) can be rewritten as

$$\begin{aligned}&\sum_{j=2}^{L}\mathbb{E}\left[\sum_{n=1}^{\Gamma_T}\mathbb{1}\left(\sum_{t=1}^{\tau_n}Z_i(t)<\frac{\alpha(\log n+\log 2)}{\Omega_{j-1,j}^2}\right)\right]\\&\leq\sum_{j=2}^{L}\sum_{n=1}^{T}\Pr\left(\sum_{t=1}^{\tau_n}Z_i(t)-p_in<\frac{\alpha(\log n+\log 2)}{\Omega_{j-1,j}^2}-p_jn\right)\\&\leq\sum_{j=2}^{L}\left(\zeta_j+\sum_{n=\zeta_j+1}^{T}\exp\left(-2t\left(p_j-\frac{\alpha(\log t+\log 2)}{t\Omega_{j-1,j}^2}\right)^2\right)\right)\\&\leq\sum_{j=1}^{L}\left(\zeta_j+\frac{1}{2p_j^2}\right). \qquad (28)\end{aligned}$$

where, $\zeta_j = \max_t\left\{\frac{p_j}{2}-\frac{\alpha(\log t+\log 2)}{t\Omega_{j-1,j}^2}<0\right\}$.

The upper bound of (24) can be obtained as

$$\sum_{j=2}^{L}\mathbb{E}\left[\sum_{n=1}^{\Gamma_T}\mathbb{1}\left(N_j(\tau_n)<\frac{\alpha\log\tau_n}{\Omega_{j-1,j}^2}\right)\mathbb{1}\left(\tau_n>2n\right)\right]$$

$$\leq \sum_{j=2}^{L} \mathbb{E}\left[\sum_{n=1}^{\Gamma_T} \mathbb{1}\left(\tau_n > 2n\right)\right]$$
$$= \sum_{j=2}^{L} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(t/2 > \sum_{s=1}^{t} \mathbb{1}(\bar{\xi}_t)\right)\right]$$
$$= \sum_{j=2}^{L} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(\sum_{s=1}^{t} \mathbb{1}(\xi_t) > t/2\right)\right]$$
$$\leq \sum_{j=2}^{L} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(\frac{\mathbb{E}[(\sum_{s=1}^{T} \mathbb{1}(\xi_s))^2]}{(t/2)^2}\right)\right] \quad (29)$$

where (29) follows Chebyshev's inequality and it can be further bounded by

$$\sum_{j=2}^{L} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(\frac{\mathbb{E}[(\sum_{s=1}^{T} \mathbb{1}(\xi_s))^2]}{(t/2)^2}\right)\right] \quad (30)$$
$$\leq \sum_{j=2}^{L} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(\frac{4}{t^2}\left(\sum_{s=1}^{t} \frac{2K}{s^2} + 2\sum_{1\leq t_1<t_2\leq t} \sqrt{\mathbb{E}[\mathbb{1}(\xi_{t_1}^2)]\mathbb{E}[\mathbb{1}(\xi_{t_2}^2)]}\right)\right)\right] \quad (31)$$
$$= \sum_{j=2}^{L} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(\frac{4}{t^2}(\sum_{s=1}^{t} \frac{2K}{s^2} + 2\sum_{1\leq t_1<t_2\leq t} \sqrt{\frac{4K^2}{t_1 t_2}})\right)\right]$$
$$= \sum_{j=2}^{L} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(\frac{8K}{t^2}(\sum_{s=1}^{t} \frac{1}{s})^2\right)\right]$$
$$< 8K\left(\frac{\log t + 1}{t}\right)^2 = \rho \quad (32)$$

where (31) is obtained by applying Cauchy's inequality to the last term.

The third part of (12) represents the case where the wrong BSs are included in $\mathcal{L}$, and these BSs in $\mathcal{L}$ are ranked in correct order. Here, we use the notation $\mathcal{K} \setminus \mathcal{L}^*$ to represent the relative complement of $\mathcal{L}^*$ with respect to $\mathcal{K}$. Let $\underline{i}_t$ be the top ranked BS from $\mathcal{K} \setminus \mathcal{L}^*$ in $\mathcal{L}_t$, and $\mathcal{D}_{i,t}$ be the event that $\underline{i}_t = i$. We have the regret of this part as

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}(\bar{\xi}_t)\mathbb{1}(\bar{\mathcal{B}}_t) reg(t)\right]$$
$$= \mathbb{E}\left[\sum_{i\in\mathcal{K}\setminus\mathcal{L}^*}\sum_{t=1}^{T} \mathbb{1}(\bar{\xi}_t)\mathbb{1}(\bar{\mathcal{B}}_t)\mathbb{1}(\mathcal{D}_{i,t})[\mathbb{1}(i\in\tilde{I}_t) + \mathbb{1}(i\notin\tilde{\mathcal{L}}_t)]reg(t)\right]$$
$$= \mathbb{E}\left[\sum_{i\in\mathcal{K}\setminus\mathcal{L}^*}\sum_{t=1}^{T} \mathbb{1}(\bar{\xi}_t)\mathbb{1}(\bar{\mathcal{B}}_t)\mathbb{1}(\mathcal{D}_{i,t})\mathbb{1}(i\in\tilde{\mathcal{L}}_t)reg(t)\right] \quad (33)$$
$$= \mathbb{E}\left[\sum_{i\in\mathcal{K}\setminus\mathcal{L}^*}\sum_{t=1}^{T} \mathbb{1}(\bar{\xi}_t)\mathbb{1}(\bar{\mathcal{B}}_t)\mathbb{1}\left(\frac{\hat{P}_i(t)+u_{i,t}}{\hat{\eta}_i(t)+u_{i,t}} > \tau\right) \cdot\mathbb{1}(i\in\tilde{I}_t)reg(t)\right]$$
$$\leq \sum_{i\in\mathcal{K}\setminus\mathcal{L}^*} K\frac{16\alpha\log(T)}{\Delta_i^2} \quad (34)$$

where (33) follows from the fact that if BS $\underline{i}_t$ is not scanned, which means $\mathcal{L}$ contains the same BSs as that in $\mathcal{L}^*$, the regret resulted from this event equals to 0. By counting the number of times that BS i is scanned, and using the fact that $\mathbb{1}\left(\frac{\hat{P}_i(t)+u_{i,t}}{\hat{\eta}_i(t)+u_{i,t}} > \tau\right) = \mathbb{1}\left(N_i(t) < \frac{16\alpha\log(T)}{\Delta_i^2}\right)$, (34) is obtained.

Combining the above results, we have the regret of proposed algorithm as

$$R(T) = \sum_{t=1}^{T}\mathbb{E}\Big[\mathbb{1}(\xi_t)reg(t)\Big] + \sum_{t=1}^{T}\mathbb{E}\Big[\mathbb{1}(\bar{\xi}_t)(\mathbb{1}(\mathcal{B}_t) + \mathbb{1}(\bar{\mathcal{B}}_t))reg(t)\Big]$$
$$\leq \rho + K(\zeta + 2\psi) + \sum_{j\in\mathcal{K}\setminus\mathcal{L}^*}\frac{16K\alpha\log(T)}{\Delta_j^2} \quad (35)$$

The upper bound in Theorem 1 is proved.

REFERENCES

[1] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K.-K. Wong, R. Schober, and L. Hanzo, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2nd Quart., 2016.

[2] *Procedures for the 5G System*, Release 16, document TS 23.502, 3GPP, 2019.

[3] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 497–508, Apr. 2012.

[4] D. K. Bhargavi, A. M. V. Prakash, "A novel handover algorithm for LTE based macro-femto heterogeneous networks," *Int. J. VLSI Des. Commun. Syst.*, vol. 6, no. 4, pp. 25–33, Aug. 2015.

[5] M. Z. Chowdhury, Y. M. Jang, and Z. J. Haas, "Network evolution and QOS provisioning for integrated femtocell/macrocell networks," 2010, *arXiv:1009.2368*. [Online]. Available: http://arxiv.org/abs/1009.2368

[6] *Evolved Universal Terrestrial Radio Access (E-UTRA), User Equipment (UE) Procedures in Idle Mode, Version 15.3.0, Release 15*, document TS 36.304, 3GPP, May 2019.

[7] Y. Watanabe, Y. Matsunaga, K. Kobayashi, H. Sugahara, and K. Hamabe, "Dynamic neighbor cell list management for handover optimization in LTE," in *Proc. IEEE 73rd Veh. Technol. Conf. (VTC Spring)*, May 2011, pp. 1–5.

[8] D. Kim, B. Shin, D. Hong, and J. Lim, "Self-configuration of neighbor cell list utilizing E-UTRAN NodeB scanning in LTE systems," in *Proc. 7th IEEE Consum. Commun. Netw. Conf.*, Jan. 2010, pp. 1–5.

[9] M. Alhabo and L. Zhang, "Load-dependent handover margin for throughput enhancement and load balancing in hetnets," *IEEE Access*, vol. 6, pp. 67718–67731, 2018.

[10] V. M. Nguyen and H. Claussen, "Efficient self-optimization of neighbour cell lists in macrocellular networks," in *Proc. 21st Annu. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 2010, pp. 1923–1928.

[11] F. Parodi, M. Kylvaja, G. Alford, J. Li, and J. Pradas, "An automatic procedure for neighbor cell list definition in cellular networks," in *Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Netw.*, Jun. 2007, pp. 1–6.

[12] M. Amirijoo, P. Frenger, F. Gunnarsson, H. Kallin, J. Moe, and K. Zetterberg, "Neighbor cell relation list and measured cell identity management in LTE," in *Proc. NOMS-IEEE Netw. Oper. Manage. Symp.*, Apr. 2008, pp. 152–159.

[13] Z. Lv, S. Yao, L. Li, Y. Qi, C. Liu, T. Li, and L. Xu, "Neighbor cell list optimization of LTE based on MR," in *Proc. Int. Conf. Signal Inf. Process., Netw. Comput.* Singapore: Springer, May 2018, pp. 290–295.

[14] K. Han, S. Woo, D. Kang, and S. Choi, "Automatic neighboring BS list generation scheme for femtocell network," in *Proc. 2nd Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Jun. 2010, pp. 251–255.

[15] M. Z. Chowdhury, M. T. Bui, and Y. M. Jang, "Neighbor cell list optimization for femtocell-to-femtocell handover in dense femtocellular networks," in *Proc. 3rd Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Jun. 2011, pp. 241–245.

[16] M. Vondra and Z. Becvar, "Distance-based neighborhood scanning for handover purposes in network with small cells," *IEEE Trans. Veh. Technol.*, vol. 65, no. 2, pp. 883–895, Feb. 2016.

[17] A. A. Gebremichail and C. Beard, "Fade duration based neighbor cell list optimization for handover in femtocell networks," *Int. J. Interdiscipl. Telecommun. Netw.*, vol. 9, no. 2, pp. 1–15, Apr. 2017.

[18] D. Soldani, G. Alford, F. Parodi, and M. Kylvaja, "An autonomic framework for self-optimizing next generation mobile networks," in *Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Netw.*, Jun. 2007, pp. 1–6.

[19] H. Zhou, "A dynamic neighbor cell list generating algorithm in cellular system," in *Proc. 5th Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Sep. 2009, pp. 1–4.

[20] Z. Becvar, M. Vondra, and P. Mach, "Dynamic optimization of neighbor cell list for femtocells," in *Proc. IEEE 77th Veh. Technol. Conf. (VTC Spring)*, Jun. 2013, pp. 1–6.

[21] Z. Becvar, P. Mach, and M. Vondra, "Self-optimizing neighbor cell list with dynamic threshold for handover purposes in networks with small cells," *Wireless Commun. Mobile Comput.*, vol. 15, no. 13, pp. 1729–1743, Sep. 2015.

[22] B. Lee, Y. Oh, J. Kim, S. Pyun, J. Moon, D. Cho, H. Lee, M. Kang, B. Jung, and J. Lee, *Scanning Overhead Reduction Based on RSSI Profiles*, Standard IEEE C802.16m-09/1399r1, Aug. 2009.

[23] A. Jain, E. Lopez-Aguilera, and I. Demirkol, "Improved handover signaling for 5G networks," in *Proc. IEEE 29th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2018, pp. 164–170.

[24] B. Kveton, C. Szepesvari, Z. Wen, and A. Ashkan, "Cascading bandits: Learning to rank in the cascade model," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 767–776.

[25] R. Zhou, C. Gan, J. Yang, and C. Shen, "Cost-aware cascading bandits," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3228–3234.

[26] R. Arshad, H. ElSawy, S. Sorour, T. Y. Al-Naffouri, and M.-S. Alouini, "Velocity-aware handover management in two-tier cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1851–1867, Mar. 2017.

[27] M. M. Hasan, S. Kwon, and S. Oh, "Frequent-handover mitigation in ultra-dense heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 1035–1040, Jan. 2019.

[28] C. Shen, C. Tekin, and M. van der Schaar, "A non-stochastic learning approach to energy efficient mobility management," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3854–3868, Dec. 2016.

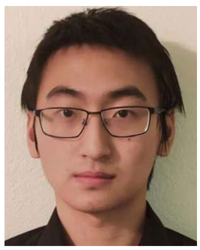
[29] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, no. 2, pp. 235–256, 2002.

[30] *Telecommunication management; Configuration Management (CM); Notification Integration Reference Point (IRP); Requirements*, document TS 32.301, 3GPP, 2015.

CHAO WANG received the B.S. degree in automation from the Department of Electrical Engineering, Hefei University of Technology, Hefei, China, in 2015. He is currently pursuing the Ph.D. degree with the Department of Automation, University of Science and Technology of China, Hefei. His research interests include reinforcement learning and wireless networks.

JIAN YANG (Senior Member, IEEE) received the B.S. and Ph.D. degrees from the University of Science and Technology of China (USTC), Hefei, China, in 2001 and 2006, respectively. From 2006 to 2008, he was a Postdoctoral Scholar with the Department of Electronic Engineering and Information Science, USTC. Since 2008, he has been an Associate Professor with the Department of Automation, USTC, where he is currently a Professor with the School of Information Science and Technology. His research interests include future networks, distributed system design, modeling and optimization, multimedia over wired and wireless networks, and stochastic optimization. He received the Lu Jia-Xi Young Talent Award from the Chinese Academy of Sciences, in 2009.

HUASEN HE received the B.S. degree in automation from the University of Science and Technology of China, Hefei, China, in 2013, and the M.S. degree in signal processing and communications and the Ph.D. degree in digital communications from The University of Edinburgh, Edinburgh, U.K., in 2014 and 2018, respectively. He is currently an Associate Research Fellow with the School of Information Science and Technology, University of Science and Technology of China. His research interests include stochastic geometry, future networks, and signal processing for communication.



RUIDA ZHOU (Graduate Student Member, IEEE) received the B.Eng. degree in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 2018. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. His research interests include information theory and statistical learning.

SHUANGWU CHEN received the B.S. and Ph.D. degrees from the University of Science and Technology of China (USTC). He is currently a Postdoctoral Researcher with USTC. His research interests include future networks, multimedia communication, and stochastic optimization.

XIAOFENG JIANG (Member, IEEE) received the B.E. and Ph.D. degrees in information science and technology from the University of Science and Technology of China (USTC), Hefei, China, in 2008 and 2013, respectively. He has been a Postdoctoral Fellow and an Associate Research Fellow with the Department of Automation, USTC, since 2013 and 2017, respectively. He has also been a Postdoctoral Fellow with the Department of Electrical Engineering, Columbia University, New York, NY, USA, since 2015. His recent research interests include future networks, 6G wireless communications, blockchain, and discrete event dynamic systems.