Context-Aware Mobility Management in HetNets: A Reinforcement Learning Approach

Meryem Simsek*, Mehdi Benni‡, and İsmail Güvenç‡

*Vodafone Chair Mobile Communications Systems Technische Universität Dresden, Germany
†Centre for Wireless Communications, University of Oulu, Finland
‡Department of Electrical & Computer Engineering, Florida International University, USA

Email: meryem.simsek@tu-dresden.de, bennis@ee.oulu.fi, and iguvenc@fiu.edu

Abstract—The use of small cell deployments in heterogeneous network (HetNet) environments is expected to be a key feature of 4G networks and beyond, and essential for providing higher user throughput and cell-edge coverage. However, due to different coverage sizes of macro and pico base stations (BSs), such a paradigm shift introduces additional requirements and challenges in dense networks. Among these challenges is the handover performance of user equipment (UEs), which will be impacted especially when high velocity UEs traverse picocells. In this paper, we propose a coordination-based and context-aware mobility management (MM) procedure for small cell networks using tools from reinforcement learning. Here, macro and pico BSs jointly learn their long-term traffic loads and optimal cell range expansion, and schedule their UEs based on their velocities and historical rates (exchanged among tiers). The proposed approach is shown to not only outperform the classical MM in terms of UE throughput, but also to enable better fairness. In average, a gain of up to 80% is achieved for UE throughput, while the handover failure probability is reduced up to a factor of three by the proposed learning based MM approaches.

Index Terms—Cell range expansion, HetNets, load balancing, mobility management, reinforcement learning, context-aware scheduling.

I. INTRODUCTION

The deployment of Long Term Evolution (LTE) heterogeneous networks (HetNets) is a promising approach to meet the ever-increasing wireless broadband capacity challenge [1], [2]. However, deploying HetNets entails a number of challenges in terms of capacity, coverage, mobility management (MM), and mobility load balancing (MLB) across multiple network tiers [3]. Mobility management is essential to ensure a continuous connectivity to mobile user equipment (UEs) while maintaining quality of service (QoS).

The mobility framework for LTE was originally developed and analyzed by the 3rd generation partnership project (3GPP) for macro-only networks, and was therefore not explicitly optimized for HetNets. In LTE Rel. 11, mobility enhancements in HetNets have been investigated through a dedicated study item [3]. 3GPP has defined key performance indicators (KPIs) for mobility measurements, i.e., the handover failure (HOF) due to a degraded signal-to-interference-plus-noise-ratio (SINR), the radio link failure (RLF), as well as the probability of unnecessary handovers, typically referred to as ping-pong (PP) events.

Poor MM approaches may increase the HOFs, RLFs, and PPs, and result in unbalanced load among cells. This entails a low resource utilization efficiency and hence deterioration of the user experience. In order to solve this problem, while minimizing PPs, mobility parameters in each cell need to be carefully and dynamically optimized according to cell traffic loads. It is essential to optimize handover parameters such as time to trigger (TTT), range expansion bias (REB), and hysteresis margin in order to answer the question: “when to handover which UE to which cell?”

Mobility management techniques for HetNets have been recently investigated in the literature, e.g., in [4]–[7]. In [4], the authors evaluate the effect of different combinations of MM parameter settings for HetNets. The main result is that mobility performance strongly depends on the cell size and UE speed. The simulations in [4] consider that all UEs have the same velocity in each simulation setup. In [5], the authors evaluate the mobility performance of HetNets considering almost blank subframes in the presence of cell range expansion and propose a mobility based intercell interference coordination (ICIC) scheme. Hereby, picocells configure coordinated resources by muting certain subframes so that macrocells can schedule their high velocity UEs in these resources without co-channel interference from picocells. However, the proposed approach only considers three broad classes of UE velocities: low, medium, and high. Moreover, no adaptation of the REB has been taken into account. In [6], a handover-aware ICIC approach based on reinforcement learning is proposed. Hereby, the authors model the ICIC approach as a sub-band selection problem for mobility robustness optimization in a small cell only network. In [7], the cell selection problem in HetNets is formulated as a network wide proportional fairness optimization problem by jointly considering the long-term channel condition and load balance in a HetNet. While the proposed method enhances the cell-edge UE performance, no results related to mobility parameters are presented.

To the best of our knowledge there is no previous work related to learning based mobility management in HetNets by jointly considering load balancing and UE scheduling. In this paper, we propose a joint MM and context-aware UE scheduling approach by using tools from reinforcement learning. Hereby, each base station (BS) individually optimizes
Our contributions are as follows:

- In the proposed MM approaches, we focus on both short-term and long-term solutions. In the long-term, a traffic load balancing procedure in a HetNet scenario is proposed, while in the short-term the UE association process is solved.
- To implement the long-term load balancing method, we propose two learning based MM approaches by using reinforcement learning techniques: a MAB based and a satisfaction based MM approach.

The short-term UE association process is based on a proposed context-aware scheduler considering a UE’s throughput history and velocity to enable fair scheduling and enhanced cell association.

The rest of the paper is organized as follows. Section II describes the system model, the problem formulation for MM, and the context-aware scheduler. In Section III, we introduce the learning based MM approaches. Section IV presents system level simulation results, and finally, Section V concludes the paper.

II. System Model

We focus on the downlink transmission of a 2-layer HetNet, where layer 1 is modeled as macrocells and layer 2 as picocells. The HetNet consists of a set of BSs $\mathcal{K} = \{1, \ldots, K\}$ with a set $\mathcal{M} = \{1, \ldots, M\}$ of macrocells underlaid by a set $\mathcal{P} = \{1, \ldots, P\}$ of picocells, where $\mathcal{K} = \mathcal{M} \cup \mathcal{P}$. Macro BSs are dropped following a hexagonal layout including three sectors. Within each macro sector $m$, $p \in \mathcal{P}$ picocells are randomly positioned, and a set $\mathcal{U} = \{1, \ldots, U\}$ of UEs which are randomly dropped within a circle around each picocell $p$ (hotspot). The UEs associated to macrocells are referred as macro UEs $\mathcal{U}(m) = \{1(m), \ldots, U(m)\}$ and the UEs served by picocells are referred as pico UEs $\mathcal{U}(p) = \{1(p), \ldots, U(p)\} \in \mathcal{U}$, where $\mathcal{U}(p) \neq \mathcal{U}(m)$.

Each UE $i(k)$ with $k \in \{m, p\}$ has a randomly selected velocity $v_i(k) \in V$ km/h and a random direction of movement within an angle of $[0; 2\pi]$. A co-channel deployment is considered, in which picocells and macrocells operate in a system with a bandwidth $B$ consisting of $r = \{1, \ldots, R\}$ resource blocks (RBs). At every time instant $t_n = nT_s$ with $n = \{1, \ldots, N\}$ and $T_s = 1$ ms, each BS $k$ decides how to expand its coverage area by learning its REB $\beta_k = \{\beta_m, \beta_p\}$ with $\beta_m = \{0; 3; 6\}$ dB and $\beta_p = \{0; 3; 6; 9; 12; 15; 18\}$ dB. Both macro and pico BSs select their REB to decide which UE $i(k)$ to schedule on which RB based on the UE’s context parameters. These context parameters are defined as the UE’s velocity $v_i(k)$, its instantaneous rate $\varphi_i(k)(t_n)$ when associated to BS $k$ and its average rate $\bar{\varphi}_i(k)(t_n)$ defined as $\bar{\varphi}_i(k)(t_n) = \frac{1}{T} \sum_{n=1}^{N} \varphi_i(k)(t_n)$, whereby $T = NT_s$ is a time window. The instantaneous rate $\varphi_i(k)(t_n)$ is given by:

$$\varphi_i(k)(t_n) = B_{i(k)} \cdot \log \left(1 + \gamma_{i(k)}(t_n)\right),$$ \hspace{1cm} (1)

with $\gamma_{i(k)}(t_n)$ being the SINR of UE $i(k)$ at time $t_n$, which is defined as:

$$\gamma_{i(k)}(t_n) = \frac{p_k \cdot g_{i(k),k}(t_n)}{\sum_{j \neq k} p_j \cdot g_{i(k),j}(t_n) + \sigma^2},$$ \hspace{1cm} (2)

with $p_k$ being the transmit power of BS $k$, and $g_{i(k),k}(t_n)$ being the channel gain from cell $k$ to UE $i(k)$ associated to BS $k$. The bandwidth $B_{i(k)}$ in equation (1) is the bandwidth which is allocated to UE $i(k)$ by BS $k$ at time $t_n$.

\footnote{We consider lower REB values for macro BSs to avoid overloaded macrocells due to their large transmission power.}
A. Handover Procedure

According to the 3GPP standard, the handover mechanism is based on RSRP measurements, the filtering of measured RSRP samples, Handover Hysteresis Margin, and TTT mechanisms [3]. A handover is executed if the target cell’s (biased) RSRP (plus hysteresis margin) is larger than the source cell’s (biased) RSRP. In summary, the handover condition for a UE $i(k)$ to BS $k$ is defined as:

$$P_i(i(l)) + \beta_i < P_k(i(k)) + \beta_k + m_{\text{hist}},$$  

with $\{l,k\} \in K$, $m_{\text{hist}}$ is the UE- or cell-specific hysteresis margin, $\beta_k(\beta_i)$ is the REB of BS $k$ (l), and $P_k(i(k))$ is the $i(k)$-th UE’s RSRP from BS $k(l)$ after TTT.

B. Problem Formulation

Our optimization approach aims at maximizing the total rate of the network. Hereby, we consider long-term and short-term processes. The long-term load balancing optimization approach is solved by the proposed learning based MM approaches presented in Section II-A and Section II-B which result in REB $\beta_k$ value optimization and in load balancing $\phi_{k,\text{to}}(t_n)$. Based on the estimated instantaneous load, the context-aware scheduler selects, in the short-term, for each RB a UE by considering its history and velocity as described in Section I-EC. This results in each UE’s instantaneous rate $\phi_i(t_n)$ and the RB allocation vector $\alpha_i(t_n)$ containing binary variables $\alpha_i(r)$, and indicating whether UE $i(k)$ to BS $k$ is allocated at RB $r$ or not. At each time instant $t_n$, each BS $k$ performs the following optimization:

$$\max_{\alpha_i(t_n)} \sum_{n=1}^{N} \sum_{k \in U_k} \sum_{r=1}^{R} \alpha_i(t_n) \cdot \phi_i(t_n),$$

subject to:

$$\alpha_i(t_n) \in \{0, 1\},$$

$$\sum_{i(k) \in U_k} \alpha_i(t_n) = 1 \ \forall r, \forall k,$$

$$p_k \leq P_k^{\text{max}},$$

$$\phi_i(t_n) \geq \phi_{k,\text{min}},$$

where $\phi_i(r,t_n)$ is the instantaneous rate of UE $i(k)$ at RB $r$. The condition in (7) implies that the total transmitted power over all RBs does not exceed the maximum transmission power $P_k^{\text{max}}$ of BS $k$.

C. Context-Aware Scheduler

The proposed MM approach does not only optimize the load according to Section II-B but considers also context-aware and fairness based UE scheduling. At each RB $r$, a UE $i(k)$ is selected to be served by BS $k$ on RB $r$ according to the following scheduling criterion:

$$i(k)_r^* = \min_{i(k) \in U_k} \left( \frac{\phi_i(r,t_n)}{\phi_i(t_n)} \right),$$

where $\text{sort}_{\text{min}}(\phi_i(t_n))$ sorts the candidate UEs according to their velocity starting with the slowest UE, i.e. if more than one UE can be selected for RB $r$, the UE with minimum velocity is selected. The rationale behind introducing the sorting/ranking function for candidate UEs according to their velocity is that high-velocity UEs will not be favored over slow moving UEs.

A scheduler according to (9) will allocate many (or even all) resources to a newly handed over UE since its average rate $\phi_{i(t_n)}(t_n)$ in the target cell is zero, i.e. in the classical Proportional Fair scheduler, $\phi_i(t_n) = \phi_i(t_n)$ when a UE is handed over to cell $k$, whereas we redefine it according to (10). To avoid this and enable a fair resource allocation among all UEs, we propose a fair based scheduling approach. We define the average rate $\phi_i(t_n)$ according to (10) incorporating the following idea: Via the X2-interface macro- and picocells coordinate, so that once a macro UE $i(m)$ is handed over to picocell $p$ its rate history at time instant $t_n$ is provided to picocell $p$ in terms of average rate $\bar{\phi}_{i(m)}(t_n)$, such that the UE’s (which is named as $i(p)$ after the handover) average rate at picocell $p$ becomes:

$$\phi_{i(p)}(t_n) = \frac{T \cdot \phi_{i(m)}(t_n) + \phi_{i(p)}(t_n + T_s)}{T + T_s}. $$

In (10), a moving average rate is considered from macrocell to picocell, whereas in the classical MM approaches a UE’s history is not considered and is equal to zero. In other words, the proposed MM approach considers the historical rate when UE $i(m)$ was associated to the macrocell $m$ in the past.

III. Learning Based Mobility Management Algorithm

To solve the optimization approach defined in Section II-B we rely on the self organizing capabilities of HetNets and propose an autonomous solution for load balancing by using tools from reinforcement learning [8]. Hereby, each cell develops its own MM strategy to perform optimal load balancing based on the proposed learning based approaches presented in Section II-A and Section II-B. To realize this, we consider the game $G = \{K, \{A_k\}_{k \in K}, \{U_k\}_{k \in K}\}$. Hereby, the set $K = \{M \cup P\}$ represents the set of players (i.e., BSs), and for all $k \in K$, the set $A_k = \{\beta_k\}$ represents the set of actions player $k$ can adopt. For all $k \in K$, the function $u_k(t_n)$ is the utility function of player $k$. The players learn at each time instant $t_n$ to optimize the load in long-term and to perform context aware scheduling in short-term based on the algorithms presented in Section II-A and II-B by the following steps:

1) Action $a_k \in A_k$ is selected based on the obtained utility $u_k(t_n) = \phi_k(t_n)$ with $\phi_{k,\text{to}}(t_n)$ being the total rate of player $k$ at time $t_n$ as defined in equation (11).

2) The action selection strategy is updated based on the selected learning algorithm presented in Section II-A and Section II-B

3) UE of BS at RB $r$ is allocated based on its velocity, its instantaneous rate, and its average rate according to (9).
A. Multi-Armed Bandit Based Learning Approach

The objective of the MAB approach is to maximize the overall system performance. MAB is a machine learning technique based on an analogy with the traditional slot machine (one armed bandit) [9]. When pulled at time $t_n$, each machine/player provides a reward. The objective is to maximize the collected reward through iterative pulls, i.e., learning iterations. The player selects its actions based on a decision function reflecting the well-known exploration-exploitation trade-off in learning algorithms.

The set of players, actions and the utility function for our MAB based MM approach is defined as follows:

- **Players**: Macro BSs $\mathcal{M} = \{1, \ldots, M\}$ and pico BSs $\mathcal{P} = \{1, \ldots, P\}$.
- **Actions**: $A_k = \{\beta_k\}$ with $\beta_m = [0, 3, 6]$ dB and $\beta_p = [0, 3, 6, 9, 12, 15, 18]$ dB being the CRE bias. We consider higher bias values for picocells due to their low transmit power. The considered bias values rely partially on the assumptions in [10] and at the same time extensive simulation results.
- **Strategy**:
  1) Every BS learns its optimum CRE bias value on a long-term basis considering its load:

$$\phi_{k, \text{tot}}(t_n) = \sum_{i(k) \in \mathcal{d}_k} \sum_{r=1}^{R} \alpha_{i(k), r}(t_n) \cdot \phi_{i(k), r}(t_n).$$

(11)

This is inter-related with the handover triggering by defining the cell border of each cell.
  2) A UE is handed over to BS $k$ if it fulfills the condition [3].
  3) RB based scheduling is performed based on equation [2].
- **Utility Function**: The utility function in MAB learning is a decision function composed by an exploitation term represented by player $k$’s total rate and exploration part considering the number of times an action has been selected so far. Player $k$ selects its action $a_{j(k)}(t_n) \in A_k$ at time $t_n$ through maximizing a decision function $d_{k, a_{j(k)}}(t_n)$, which is defined as:

$$d_{k, a_{j(k)}}(t_n) = u_{k, a_{j(k)}}(t_n) + \frac{2 \log \left( \sum_{i=1}^{|A_k|} n_{k, a_{i(k)}}(t_n) \right)}{n_{k, a_{j(k)}}(t_n)}.$$  

(12)

whereby $u_{k, a_{j(k)}}(t_n)$ is the mean reward of player $k$ at time $t_n$ for action $a_{j(k)}$. $n_{k, a_{j(k)}}(t_n)$ is the number of times action $a_{j(k)}$ has been selected by player $k$ until time $t_n$, and $|\cdot|$ represents the cardinality.

During the first $t_n = |A_k| \cdot T_s$ player $k$ selects each action once in a random order to initialize the learning process by receiving a reward for each action. For the following iterations $t_n > |A_k| \cdot T_s$ action selection is performed according to Algorithm 1. In each learning iteration the action $a^{*}_{j(k)}$ that maximizes the decision function in (12) is selected. Then the parameters are updated, whereby the following notation is used: $s_{k, a_{j(k)}}(t_n)$ is the cumulated reward of player $k$ after playing action $a_{j(k)}$ and $\mathbb{1}_{i=j}$ is equal to 1 if $i = j$ and zero otherwise.

B. Satisfaction Based Learning Approach

Satisfaction based learning approaches guarantee to satisfy the players in a system [11]. Here, we consider the player to be satisfied if its cell reaches a certain minimum level of total rate and if at least 90% of the UEs in the cell obtain a certain average rate. The rationale behind considering these satisfaction conditions is to guarantee each single UE’s minimum rate while at the same time improving the total rate of the cell.

To enable a fair comparison, the set of players and the corresponding set of actions in the proposed satisfaction based MM approach are the same as in the MAB based MM approach. The utility function of player $k$ at time $t_n$ is defined as the load according to equation (11). In the satisfaction based learning approach, the actions are selected according to a probability distribution $\pi_{k}(t_n) = [\pi_{k,1}(t_n), \ldots, \pi_{k,|A_k|}(t_n)]$. Hereby, $\pi_{k,j}(t_n)$ is the probability with which BS $k$ chooses its action $a_{j(k)}(t_n)$ at time $t_n$. The following learning steps are performed in each learning iteration:

1) In the first learning iteration $t_n = 1$ the probability of each action is equal and an action is selected randomly.
2) In the following learning iterations $t_n > 1$, the player changes its action selection strategy only if the received utility does not satisfy the cell, i.e., if the satisfaction condition is not fulfilled.
3) If the satisfaction condition is not fulfilled, the player $k$ selects its action $a_{j(k)}(t_n)$ according to the probability distribution $\pi_{k}(t_n)$.
4) Each player $k$ receives a reward $\phi_{k, \text{tot}}(t_n)$ based on the selected actions.
5) The probability $\pi_{k,j}(t_n)$ of action $a_{j(k)}(t_n)$ is updated according to the linear reward-in-action scheme:

$$\pi_{k,j}(t_n) = \pi_{k,j}(t_n - T_s) + \lambda \cdot b_k(t_n),$$

(13)

whereby $b_k(t_n) = 1$ for the selected action and zero for the non-selected actions and $b_k(t_n)$ is defined as follows:

Algorithm 1 MAB based mobility management algorithm.

1: for $t_n$ do
2: for $i = 1 : |A_k|$ do
3: Select action $a_{j(k)}$ according:
4: $a_{j(k)} = \arg \max_{a_{j(k)} \in A_k} \left( d_{k, a_{j(k)}}(t_n) \right)$
5: Update parameters according to:
6: Update the cumulated reward when player $k$ selects action $a_{j(k)}$
7: $s_{k, a_{j(k)}}(t_n + T_s) = s_{k, a_{j(k)}}(t_n) + \mathbb{1}_{i=j} \cdot \phi_{k, \text{tot}}(t_n)$
8: $n_{k, a_{j(k)}}(t_n + T_s) = n_{k, a_{j(k)}}(t_n) + \mathbb{1}_{i=j}$
9: $u_{k, a_{j(k)}}(t_n + T_s) = s_{k, a_{j(k)}}(t_n + T_s) / n_{k, a_{j(k)}}(t_n + T_s)$
10: end for
11: $t_n = t_n + T_s$
12: end for
macro and pico BSs. This baseline approach is referred to as fair based scheduling, with no information exchange between TTT. For the baseline MM approach, we consider proportional condition in (3). A UE is handed over to its target cell after filtered handover measurement meets the handover event entry as a Layer 3 filter. A handover is then triggered if the Layer 3 out fast fading effects. This value is further averaged through averages the reported RSRP values every 200 ms to filter every 40 ms and reports this value. The Layer 1 filtering [3]. The UE performs RSRP measurements over one subframe computation times due to the complexity of our system level simulations. Each UE is performed with the picocell deployment based modified version of the system level simulator presented in [13]. We consider a macrocell consisting of three sectors, an inter-side distance of 500 m, and $P = \{1, 2, 3\}$ pico BSs per macro sector, randomly distributed within the macrocellular environment. In each macro sector, $U = 30$ mobile UEs are randomly dropped within a 60 m radius of each pico BS. The rationale behind dropping all UEs around pico BSs is to obtain a large number of handover within a short time in order to avoid large computation times due to the complexity of our system level simulations. Each UE $i(k)$ has a randomly selected velocity $v_{i(k)}$ of $V = \{3; 30; 60; 120\}$ km/h and a random direction of movement within $[0; 2\pi]$, so that both macro-to-pico and pico-to-macro handover may occur. We consider fast-fading and shadowing effects in our simulations that are based on 3GPP assumptions [12]. To compare our results with other approaches we consider a baseline MM approach as defined in [3]. The UE performs RSRP measurements over one subframe every 40 ms and reports this value. The Layer 1 filtering averages the reported RSRP values every 200 ms to filter out fast fading effects. This value is further averaged through a first-order infinite impulse response (IIR-)filter which is known as a Layer 3 filter. A handover is then triggered if the Layer 3 filtered handover measurement meets the handover event entry condition in [3]. A UE is handed over to its target cell after TTT. For the baseline MM approach, we consider proportional fair based scheduling, with no information exchange between macro and pico BSs. This baseline approach is referred to as classical HO approach.

$$b_k(t_n) = \frac{u_{k,\max} + \hat{u}_{k,\text{tot}}(t_n) - u_{k,\min}}{2 \cdot u_{k,\max}}$$

(14)

with $u_{k,\max}$ being the maximum rate in case of single-UE and $u_{k,\min} = \frac{1}{100+T_w}$ is the learning rate.

IV. SIMULATION RESULTS

The scenario used in the system-level simulations is based on configuration #4b HetNet scenario in [12]. Simulations are performed with the picocell deployment based modified version of the system level simulator presented in [13]. We consider a macrocell consisting of three sectors, an inter-side distance of 500 m, and $P = \{1, 2, 3\}$ pico BSs per macro sector, randomly distributed within the macrocellular environment. In each macro sector, $U = 30$ mobile UEs are randomly dropped within a 60 m radius of each pico BS. The rationale behind dropping all UEs around pico BSs is to obtain a large number of handover within a short time in order to avoid large computation times due to the complexity of our system level simulations. Each UE $i(k)$ has a randomly selected velocity $v_{i(k)}$ of $V = \{3; 30; 60; 120\}$ km/h and a random direction of movement within $[0; 2\pi]$, so that both macro-to-pico and pico-to-macro handover may occur. We consider fast-fading and shadowing effects in our simulations that are based on 3GPP assumptions [12]. To compare our results with other approaches we consider a baseline MM approach as defined in [3]. The UE performs RSRP measurements over one subframe every 40 ms and reports this value. The Layer 1 filtering averages the reported RSRP values every 200 ms to filter out fast fading effects. This value is further averaged through a first-order infinite impulse response (IIR-)filter which is known as a Layer 3 filter. A handover is then triggered if the Layer 3 filtered handover measurement meets the handover event entry condition in [3]. A UE is handed over to its target cell after TTT. For the baseline MM approach, we consider proportional fair based scheduling, with no information exchange between macro and pico BSs. This baseline approach is referred to as classical HO approach.

Fig. 2: CDF of the UE throughput for 30 UEs and 1 pico BS per macrocell and TTT = 480 ms.

$$\text{Cell-edge UE throughput}$$

$$\text{Sum-rate [Mbps]}$$

Fig. 3: Sum-rate vs. number of pico BSs per macrocell and TTT = 480 ms.

Fig. 4 depicts the cumulative distribution function (CDF) of the UE throughput for the classical, MAB and satisfaction based MM approaches. Compared to the classical approach, MAB and satisfaction based approaches lead to an improvement of 43% and 75% in average (50-th %), respectively. Hence, the satisfaction based approach outperforms the other MM approaches in terms of average UE throughput. In case of the cell-center UE throughput, which is defined as the 95-th % throughput, the opposite behavior is obtained. In this case an improvement of 124% and 80% is achieved for the MAB and satisfaction based approaches, respectively. The reason is that the satisfaction based MM approach only aims at satisfying the network in terms of rate and does not update its learning strategy once satisfaction is achieved. The MAB based approach on the other hand aims at maximizing the network performance, which is reflected in the improved cell-center UE throughput. The gains of the proposed MM approaches are also reflected in the cell-edge UE throughput, which is zoomed in Fig. 2. Here, the MAB and satisfaction based approaches yield 39% and 80% improvement compared to the classical approach.

To compare the performance of the proposed approaches for different number of picocells per macrocell, Fig. 5 plots the sum-rate versus number of pico BSs per macrocell. For different number of pico BSs the proposed MM approaches yield gains of around 70%-80 % for TTT = 480 ms. In Fig. 6 the performance of the sum-rate versus UE density per macrocell is depicted for TTT = 40 ms and TTT = 480 ms. In both cases, the classical approach yields very low rates, while the proposed approaches lead to significant improvement of up to 81 % for TTT = 40 ms and 85 % for TTT = 480 ms and converge to a significantly larger sum-rate than the classical approach.

Besides the gains in terms of rate, our proposed learning based approaches yield also improvements in terms of HOF probability as depicted in Fig. 5. For the HOF performance evaluation, we modify our simulation settings by setting the same velocity for each UE. Compared to the classical MM approach, the proposed methods yield the same HOF
Fig. 4: Sum-rate vs. number of UEs per macrocell with 1 pico BS and TTT = 40 ms and TTT = 480 ms.

Fig. 5: HOF and ping pong probability for 30 UEs and 1 pico BS per macrocell and TTT = 40 ms and TTT = 480 ms.

Fig. 6: PP probability for 30 UEs and 1 pico BS per macrocell and TTT = 40 ms and TTT = 480 ms.

probability for UEs at 3 km/h speed. For higher velocities in which more HOF is expected, the HOF probability obtained by the proposed approaches is significantly lower than in case of classical MM.

The PP probability is depicted in Fig. 6. For TTT = 40 ms, all MM methods yield very similar PP probabilities for lower velocities while this probability is decreased for higher velocities. This slope is aligned with the results presented in [3]. However, for high velocity UEs, the PP probability of the proposed MM approaches is half of the PP probability obtained for the classical MM approach which shows a significant improvement. The rationale behind this is that both tiers perform CRE for load balancing, i.e. if one cell tries to extend its coverage/handover a UE the other cell may prevent this handover by extending its coverage, too. In case of TTT = 480 ms almost no PPs are observed.

V. CONCLUSION

We propose two learning based MM approaches and a history based context-aware scheduling method for HetNets. The first learning approach is based on MAB methods and aims at system performance maximization. The second learning method aims at satisfying each cell and each UE of a cell based on satisfaction based learning. System level simulations demonstrate the performance enhancement of the proposed approaches compared to the classical MM method. While up to 80% gains are achieved in average for UE throughput, the HOF probability is reduced up to a factor of three by the proposed learning based MM approaches.

REFERENCES

[1] “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013-2018,” Cisco Public Information, February, 2014.
[2] A. Damnjanovic, et al., “A Survey on 3GPP Heterogeneous Networks”, IEEE Wirel. Comm., vol.18, no. 3, June 2011.
[3] 3GPP TR 36.839, “Evolved Universal Terrestrial Radio Access (E-UTRA); Mobility Enhancements in Heterogeneous Networks,” V11.1.0, 2012.
[4] S. Barbera, P. H. Michaelsen, M. Såd'ily, and K. Pedersen, “Mobility Performance of LTE Co-Channel Deployment of Macro and Pico Cells,” in Proc. IEEE Wireless Comm. and Networking Conf. (WCNC), France, Apr. 2012.
[5] D. Lopez-Perez, I. Guvenc, and X. Chu, “Mobility Management Challenges in 3GPP Heterogeneous Networks,” IEEE Comm. Mag., vol. 50, no. 12, Dec. 2012.
[6] A. Feki, V. Capdevielle, L. Roulet, and A.G.Sanches, “Handover aware interference management in LTE small cells networks,” in Proc. IEEE 11th International Symposium on Modeling & Optimization in Mobile, Ad Hoc & Wireless Networks (WiOpt), May 2013.
[7] J. Wang, J. Liu, D. Wang, J. Pang, and G. Shen, “Optimized Fairness Cell Selection for 3GPP LTE-A Macro-Pico HetNets,” in Proc. IEEE Vehic. Technol. Conf. (VTC), CA, Sept. 2011.
[8] M. E. Harmon and S. S. Harmon, “Reinforcement Learning: A Tutorial,” 2000. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.33.2480&rep=rep1&type=pdf
[9] P. Auer, N. Cesa-Bianchi and P. Fischer, “Finite time Analysis for the Multiarmed Bandit Problem”, Machine Learning, vol. 47, pp. 235-256, 2002.
[10] 3GPP R1-113806, “Performance Study on ABS with Reduced Macro Power,” 3GPP TSG-RAN Meeting #67, Nov. 2011.
[11] S. Ross and B. Chai-t-draa, “Satisfaction Equilibrium: Achieving Cooperation in Incomplete Information Games,” 19th Canadian Conf. on Artificial Intelligence, 2006.
[12] 3GPP TR 36.814, “Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA Physical Layer Aspects,” V9.0.0, 2010.
[13] M. Simsek, T. Akbudak, B. Zhao, and A. Czyliwicz, “An LTE-Femtocell Dynamic System Level Simulator,” International ITG Workshop on Smart Antennas (WSA), 2010.