Learn to Adapt: Self-Optimizing Small Cell Transmit Power with Correlated Bandit Learning

Zhiyang Wang*, Cong Shen*, Xiliang Luo†, Mihaela van der Schaar‡§
*University of Science and Technology of China
†ShanghaiTech University
‡University of California, Los Angeles
§University of Oxford

***Abstract*—Judiciously setting the base station transmit power that matches its deployment environment is a key problem in ultra dense networks and heterogeneous in-building cellular deployments. A unique characteristic of this problem is the tradeoff between sufficient *indoor coverage* and limited *outdoor leakage*, which has to be met without explicit knowledge of the environment. In this paper, we address the small base station (SBS) transmit power assignment problem based on stochastic bandit theory. We explicitly consider power switching penalties to discourage frequent changes of the transmit power, which causes varying coverage and uneven user experience. Unlike existing solutions that rely on RF surveys in the target area, we take advantage of the user behavior with simple coverage feedback in the network. In addition, the proposed power assignment algorithms follow the Bayesian principle to utilize the available *prior knowledge* and *correlation structure* from the self configuration phase. Simulations mimicking practical deployments are performed for both single and multiple SBS scenarios, and the resulting power settings are compared to the state-of-the-art solutions. Significant performance gains of the proposed algorithms are observed.**

I. Introduction

The massive deployment of low-power low-cost small base stations (SBS) has been viewed as an important solution to address the challenge of exponential growth of the wireless data traffic, particularly for indoor users [1]. In practice, SBSs may be deployed in drastically different scenarios, from large warehouses to small residential apartments and single-office enterprises. Correspondingly, the transmit power that determines the coverage range cannot be the same but must be decided based on the individual deployment. Furthermore, indoor enterprise deployments often have stringent access and security constraints. As a result, judiciously setting the SBS transmit power to automatically match the deployment environment is among the most important challenges for in-building SBS network deployment [2].

Several solutions for SBS transmit power self-optimization have already been proposed under the general framework of self-organizing networks (SON) [3]. Small Cell Forum has defined a common network monitor mode [4], allowing SBS to periodically measure the RF conditions and adjust its transmit power. This solution is coarse and may cause RF mismatch. In [5], a heuristic solution was proposed to determine the coverage based on RF survey. The solution has some adaptability but still lacks accuracy. The authors of [6] and [7] modeled dynamic SBS power management as a reinforcement learning problem. The main objective, however, is to adjust the transmit power in reaction to the fast-changing circumstance, which makes it more of a *power control* problem that has to be solved at a fast time scale.

This work has been supported by the National Natural Science Foundation of China under Grant 61631017.

In this work, we focus on setting the SBS transmit power of an enterprise network in an unknown environment. We limit our attention to the *closed access* mode, which is commonly adopted in enterprise networks due to security and management considerations. An adequate power assignment is particularly crucial for this setting, as the power needs to be large enough to provide sufficient coverage to the enterprise users while small enough to not create significant interference for the non-enterprise co-channel users.

Due to the unknown environment, a good solution must complement the performance *optimization* problem with an *online learning* approach to remove the uncertainty. The SBSs have to balance the immediate gains (selecting a power level that performs the best so far) and long-term performance (evaluating other options). We thus resort to multi-armed bandit (MAB) theory [8] to address the resulting exploration and exploitation tradeoff. However, as opposed to directly applying classical MAB algorithms such as UCB [9], we leverage three unique characteristics of practical SBS networks and develop a novel algorithm. First, SBS power assignment falls into the *self-optimization* category of SON, which typically follows a *self-configuration* phase that already generates some *prior knowledge* of the system. Second, in practice, performances of similar power levels are often very similar, which means that if we adopt the MAB model, nearby arms are highly *correlated*. Intuitively, such correlation can be used to accelerate the convergence to the optimal selection. Lastly, practical deployment often wants to avoid frequent power switchings, because it may cause frequent change of the coverage area and result in uneven user experience. We thus explicitly penalizes power change by adding a switching cost to the objective function, thus discouraging frequent change of power levels in the developed algorithm. These characteristics were not fully utilized in classical MAB solutions of [8], [9], and has not been utilized in wireless networks [6], [7], [10].

In this paper, we exploit these characteristics and develop a novel algorithm to solve the enterprise power assignment problem. We design a Bayesian learning [11] based algorithm that simultaneously incorporates both the prior estimates from the self-configuration phase and correlation [12] among different power levels that captures the similarity among nearby power levels. Furthermore, we explicitly add a switching cost to the performance function in order to discourage frequent change of power levels, and develop a *block allocation* scheme that combines the effect of switching cost and Bayesian learning. For the proposed Correlated Bayesian Power Assignment with Switching Cost (CBPA-SC) algorithm, rigorous analysis of the performance loss with respect to the genie-aided global optimization solution is carried out. Furthermore, we consider the complexity issue associated with CBPA-SC in a multi-SBS deployment, and introduce a *clustering* based solution that utilizes the prior knowledge of the power levels. The performance gain of CBPA-SC is verified by extensive system simulations. In addition, the superior performance comes with only one bit feedback per user location, as opposed to the full-blown RF feedback that is required in the existing solutions.

The rest of the paper is organized as follows. The system model is given in Section II. Section III presents the proposed power assignment algorithm, with performance and complexity analysis. Simulation results are presented in Section IV. Finally, Section V concludes the paper.

II. System Model

Both single-SBS and multi-SBS deployments are considered. Note that the former is suitable for modeling single-office enterprises and other small deployment, while the latter mainly applies to large warehouses, for which multiple SBSs are installed to jointly cover the indoor users. The set of SBSs is indexed as $\mathcal{K}_{SBS} = \{1, 2, .., K\}$. Each SBS has a set of candidate pilot power levels[1], denoted as $\mathcal{P} = \{p_1, p_2, .., p_n\}$. We assume that the users at measurement points inside the enterprise building are served by the SBS network, while users at points outside can only be served by one of the MBSs from $\mathcal{K}_{MBS} = \{1, 2, .., K_M\}$, as Fig. 1 depicts. Our work only requires UE to feed back whether it is covered at a location. For non-enterprise UEs, we rely on the *registration attempt* to determine the coverage events [5]. In this work, our model and procedure on power assignment follow the common industry SON operations [3]. Specifically, the power assignment policy is executed at the central network controller.

We focus on the pilot power assignment problem where the pilot power remains stable for a relatively long period of time (e.g. minutes to hours). To formulate the problem, we first denote the set of measurement points on the inside and outside routes as $N_{in} = \{1, 2, .., n_{in}\}$ and $N_{out} = \{1, 2, .., n_{out}\}$,

[1]As the purpose of the long-term power assignment is to determine the appropriate coverage that fits the deployment, we focus on setting the pilot power instead of the power of data and control channels.

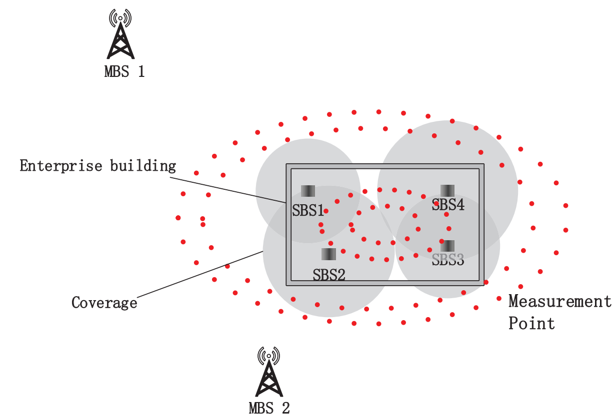

Fig. 1. An exemplary deployment of a multi-SBS setting.

respectively. The coverage and leakage criteria for a measurement point can be formally defined as:

$$\text{coverage:} \quad \max_{k_S \in \mathcal{K}_{SBS}} \mathsf{SINR}_{\mathsf{k_S,n}} > \mathsf{SINR}_{\mathsf{th}}, \text{for } n \in N_{in}, \quad (1)$$

$$\text{leakage:} \quad \max_{k_M \in \mathcal{K}_{MBS}} \mathsf{SINR}_{\mathsf{k_M,n}} < \mathsf{SINR}_{\mathsf{th}}, \text{for } n \in N_{out}, \quad (2)$$

where $\mathsf{SINR}_{\mathsf{k_S,n}}$ and $\mathsf{SINR}_{\mathsf{k_M,n}}$ represent the SINR of the measurement point n inside served by SBS k_S and outside served by MBS k_M, respectively. N_s denotes the noise and interference, and $\mathsf{SINR}_{\mathsf{th}}$ is the SINR threshold. The overall system coverage and leakage are defined as the percentages η_{in}, η_{out} of points satisfying the coverage condition (1) and leakage condition (2), respectively. Note that a larger pilot transmit power may simultaneously increase the coverage and leakage percentage. Hence, the system performance indication function (PIF) associated with each candidate power level must balance coverage and leakage. Although any meaningful PIFs that capture the tradeoff between coverage and leakage can be used, we adopt a simple linear PIF in this work as

$$r = \alpha\eta_{in} - (1 - \alpha)\eta_{out}, \quad (3)$$

where α is a parameter controlling the weight between coverage and leakage. Strictly speaking, the function r in (3) is a random variable for a given pilot power level due to the random channel effect such as shadowing, fast fading and other disturbance. We focus on a probabilistic model with *Gaussian* random fluctuation around the mean. It is worth noting that Gaussian process is a powerful and commonly used tool, as it is generally accepted as the most flexible and captures prediction under various uncertainty information [13].

III. Power Assignment Algorithm Description

A. Stochastic Bandit Model

We take a stochastic MAB approach to balance the short-term and long-term performances. Specifically, we model the set of candidate pilot power values $\mathcal{P} = \{p_1, p_2, .., p_n\}$ as n arms, denoted by $\mathcal{N}_{pow} = \{1, 2, .., n\}$. At the beginning of each time slot $t = 1, 2, .., T$, a power value $p_{a(t)} \in \mathcal{P}, a(t) \in$

$\mathcal{N}_{pow}$ is selected. At the end of the time slot t, the SBS observes a PIF feedback $r_{a(t)}(t)$ based on UE measurement reports, corresponding to $reward$ in the bandit theory. In practice, switching the power value of SBS to another value may lead to additional performance loss, e.g. interference and user experience degradation like increasing dropped call rate. To address this problem, we explicitly add a switching cost when the power level changes. For the multi-SBS case, each arm corresponds to a set of power levels of all K SBSs, and switching cost is the sum of individual costs of all SBSs while other definitions remain the same.

We adopt a general switching loss function $s_{ij} = f(|p_i - p_j|)$, a non-decreasing function of the difference between the two power values. s_{ij} is incurred whenever SBS changes its pilot power value between p_j and p_i. Therefore, the actual cumulative PIF excluding the switching cost up to a given time horizon $T > 0$ can be denoted formally as:

$$G_T^S = G_T - \mathsf{SC}(T) = \sum_{t=1}^{T} r_{a(t)}(t) - \sum_{t=2}^{T} s_{a(t)a(t-1)}. \quad (4)$$

In multi-armed bandit theory, *expected cumulative regret* [8] is often used to characterize the performance, representing the cumulative difference between the reward of the selected arms and the maximum expected reward. We comment that minimizing regret is equivalent to maximizing cumulative reward and the regret is equivalent to the performance loss of any power assignment problem *due to learning*. Therefore, our objective is to develop an efficient power assignment solution to minimize the expected cumulative PIF loss with minimum switches. We define the cumulative PIF loss as

$$R_T = G_T^* - G_T^S = \max_{i \in \mathcal{N}_{pow}} \left(\sum_{t=1}^{T} r_i(t) \right) - \sum_{t=1}^{T} r_{a(t)}(t) + \mathsf{SC}(t). \quad (5)$$

Here the optimal power level can be obtained by a genie-aided solution, e.g. a global optimization of the expected PIF with complete RF information from the technician survey. Our focus is the PIF loss of the system R_T (5) for any given time horizon T. The expected PIF loss can be written as:

$$\mathbb{E}[R_T] = T\mu^* - \mathbb{E} \sum_{t=1}^{T} \mu_{a(t)} = \sum_{i=1}^{n} \Delta_i \mathbb{E}[N_i(T)], \quad (6)$$

where $\mu^* = \max_{i \in \mathcal{N}_{pow}} \mu_i$ is the true mean PIF of the optimal power level and $\Delta_i = \mu^* - \mu_i$ measures the mean PIF gap between the chosen power level and the optimum. $N_i(T)$ represents the number of times power level p_i is selected. According to the ground-breaking work of Lai and Robbins [14], if the expected loss $\mathbb{E}[R_T]$ of our proposed algorithms can be upper bounded by $\mathcal{O}(\log T)$, an asymptotically optimal performance is achieved in the sense that the convergence rate is of the same order as the optimum.

B. Correlated Bayesian Power Assignment Algorithm

Our algorithm utilizes the *prior* knowledge of the PIF *before* the algorithm is invoked. In practice, the most common form for the prior knowledge comes from the self-configuration phase of SON, which is performed during network initialization. The self-configuration phase can provide us some prior estimations and structure of the PIFs as it typically tries different power levels before settling on one. The prior knowledge we import involves PIF estimations and the correlation structure. The PIFs of similar transmit power levels are generally correlated due to the slow and continuous changing nature of RF propagation. It is worth noting that the proposed algorithm also works with none or part of prior knowledge, at the expense of slower convergence.

We adopt the well-known Bayesian principle [11] that integrates the prior distribution and quantiles of the posterior distribution. Let $\mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0)$ be a correlated prior assumption while Σ_0 is a positive definite matrix capturing the correlation structure. We define $\{\boldsymbol{\phi}_t \in \mathbb{R}^N\}_{t \in \{1,..,T\}}$ as the indicator vector to reveal the currently selected power value $p_{a(t)}$, i.e.,

$$(\boldsymbol{\phi}_t)_k = \begin{cases} 1 & k = a(t), \\ 0 & \text{otherwise}. \end{cases} \quad (7)$$

The estimation of the mean PIFs and correlation of the PIF $(\boldsymbol{\mu}_t, \Sigma_t)$ is updated following the Bayesian principle [15]:

$$\begin{aligned} \mathbf{q}_t &= \frac{r_t \boldsymbol{\phi}_t}{\sigma_0^2} + \hat{\Lambda}_{t-1}\hat{\boldsymbol{\mu}}_{t-1}, \quad & \hat{\Lambda}_t &= \frac{\boldsymbol{\phi}_t \boldsymbol{\phi}_t^T}{\sigma_0^2} + \hat{\Lambda}_{t-1}, \\ \hat{\Sigma}_t &= \hat{\Lambda}_t^{-1}, & \hat{\boldsymbol{\mu}}_t &= \hat{\Sigma}_t \mathbf{q}_t = \hat{\Lambda}_t^{-1}\mathbf{q}_t, \end{aligned} \quad (8)$$

where r_t is the PIF observed at time slot t. To derive a general expression of the estimation, we introduce a diagonal matrix $P(t)$ with entries $\sigma_0^2/N_i(t), i \in \mathcal{N}_{pow}$, and $\bar{\mathbf{r}}_t$ is the vector of $\bar{r}_i(t), i \in \mathcal{N}_{pow}$. We first rewrite the expression of $\hat{\Lambda}_t$ as:

$$\hat{\Lambda}_t = \frac{\boldsymbol{\phi}_t \boldsymbol{\phi}_t^T}{\sigma_0^2} + \ldots + \frac{\boldsymbol{\phi}_1 \boldsymbol{\phi}_1^T}{\sigma_0^2} + \Lambda_0 = P(t)^{-1} + \Lambda_0. \quad (9)$$

Then, $\hat{\boldsymbol{\mu}}_t$ can be derived based on (9) as

$$\begin{aligned} \hat{\boldsymbol{\mu}}_t &= \hat{\Lambda}_t^{-1} \left(\frac{r_t \boldsymbol{\phi}_t}{\sigma_0^2} + \frac{r_{t-1}\boldsymbol{\phi}_{t-1}}{\sigma_0^2} + \ldots + \frac{r_1 \boldsymbol{\phi}_1}{\sigma_0^2} + \Lambda_0 \boldsymbol{\mu}_0 \right) \\ &= (\Lambda_0 + P(t))^{-1}(P(t)^{-1}\bar{\mathbf{r}}_t + \Lambda_0 \boldsymbol{\mu}_0). \end{aligned} \quad (10)$$

Therefore, the estimation at time slot t can be derived combining equation (9) and (10).

We propose a *block allocation* scheme to address the switching costs. Block allocation schemes determine specific intervals of time over which the selection is consistent. The construction of the intervals should control the expected number of switches to guarantee good performance [16]. The idea is graphically presented in Fig. 2. We first divide time into frames whose last time slot is denoted as $L_f, f \in \{1, 2...l\}, l = \lceil \sqrt{\log_2 T} \rceil$. Each frame is then subdivided into $b_f = \lceil (2^{f^2} - 2^{(f-1)^2})/f \rceil$ blocks each of which contains f time slots. Each block is identified by $(f, k), f \in \{1, 2, .., l\}, k \in \{1, 2, .., b_f\}$, with f and k representing the frame number and block number within the frame respectively. The beginning time slot of block k in the f-th frame is denoted as τ_{fk}.

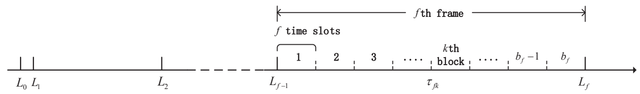

Fig. 2. The block allocation scheme in CBPA-SC.

Algorithm 1 The CBPA-SC Algorithm

Input: Prior estimation of PIF mean: $\mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0)$;

Initialize: $N_i, \bar{r}_i, Q_i = 0, \hat{\mu}_i = \mu_i^0, \hat{\Sigma} = \Sigma_0$ for all $i \in \mathcal{N}_{pow}$,

1: **for** $f \in \{1, 2, .., l\}$ **do**

2: **for** $k \in \{1, 2, .., b_f\}$ **do**

3: Set $\tau_{fk} = L_{f-1} + 1 + f(k-1)$;

4: Update $Q_i = \hat{\mu}_i + \hat{\sigma}_i \sqrt{\sum_{j=1}^{n} \rho_{ij}^2 \Phi^{-1}(1 - 1/(\sqrt{2\pi e}\tau_{fk}^2))}$ for each $i \in \mathcal{N}_{pow}$;

5: Determine $a^* = \arg\max\{Q_i | i \in \mathcal{N}_{pow}\}$;

6: Use power value p_{a^*} for the next $(n_f - 1)$ slots;

7: Collect $r_{a^*}(t)$ without the switching loss $s_{a(t)a(t-1)}$ for each $t \in \{\tau_{fk}, \tau_{fk}+1, .., T_e\}$, $T_e = \tau_{fk} + f - 1, n_f = f$ if $\tau_{fk} + f - 1 \leqslant T$, otherwise $T_e = T, n_f = T - \tau_{fk} + 1$;

8: Update the following:
$\bar{r}_{a^*} = \frac{N_{a^*}\bar{r}_{a^*} + \sum_{t=\tau_{fk}}^{T_e}(r_{a^*}(t) - s_{a(t)a(t-1)})}{N_{a^*} + n_f}$,
$N_{a^*} = N_{a^*} + n_f$,
$\hat{\boldsymbol{\mu}} = (\Sigma_0^{-1} + P^{-1})^{-1}(P^{-1}\bar{\mathbf{r}} + \Sigma_0^{-1}\boldsymbol{\mu}_0)$,
$\hat{\Sigma}^{-1} = \Sigma_0^{-1} + P^{-1}$.

9: **end for**

10: **end for**

The proposed Correlated Bayesian Power Assignment with Switching Cost (CBPA-SC) algorithm is given in Algorithm 1. CBPA-SC is based on three key ideas. The first is that the utility function defined in step 4 is composed of an estimated performance term and a measure of uncertainty, which reflects the tradeoff between exploration and exploitation. More specifically, $\Phi^{-1} : (0, 1) \to \mathbb{R}$ is the inverse cumulative distribution function (CDF) for a standard Gaussian random variable. $\mathbb{P}(\mu_i \leq Q_i(t)) = 1 - 1/(\sqrt{2\pi e}t^2)$ indicates that the true mean PIF μ_i is more likely to be less than the estimation Q_i as time goes by, leading to the convergence to the optimal power level. The second is that the correlation structure accelerates the convergence. Intuitively, if a transmit power level results in a bad PIF, then the algorithm does not need to waste much exploration on the nearby power levels, as they are likely to be bad as well. Finally, since the switching cost results in a penalty in performance, the algorithm needs to "explore in bulk". The block size increases with time to take advantage of the better knowledge about the optimal power level.

C. Reducing Complexity in The Multi-SBS Deployment

Algorithm 1 in a multi-SBS deployment will have a complexity that grows exponentially with K. To reduce the complexity, we first explore a practical constraint that the neighboring SBSs are generally not allowed to have vastly different pilot power levels. This is because otherwise they result in significantly different coverage areas and lead to uneven load distributions. Thus, we should only consider the combinations in which neighboring SBS power levels are different by no more than a certain threshold P_{th}.

Secondly, we note that for two power settings that differ only slightly, the performances may be very similar. Thus, if we can carefully group the power settings into a few *clusters*, and only use the cluster center as the representative power, we can achieve a good tradeoff between complexity and performance for the algorithms. We adopt the *K-medoids* clustering [17] based on the most central object. The choice of the number of clusters N plays a critical role in the overall performance which will be shown in Sec.IV.

D. Performance Analysis

To analyze the performance of CBPA-SC theoretically, we focus on the cumulative PIF loss given in (5) and analyze the convergence speed. Theorem 1 guarantees that the cumulative PIF of CBPA-SC will converge to that of the global optimum power value at a rate of $\mathcal{O}(\log T/T)$. Moreover, this upper bound applies to any finite time T and any switching loss function $f(|p_i - p_j|)$ as long as f is non-decreasing finite.

Theorem 1. *The expected cumulative PIF loss $\mathbb{E}[R_T^{SC}]$ of CBPA-SC is bounded above as:*

$$\begin{aligned}
\mathbb{E}[R_T^{SC}] &\leqslant \sum_{i=1, i\neq i^*}^{n} \Delta_i \mathbb{E}[N_i(T)] + \mathbb{E}[\mathsf{SC}(t)] \\
&\leqslant \sum_{i=1, i\neq i^*}^{n} \Delta_i (C_1^i \log T + C_2^i) + \sum_{i=1, i\neq i^*}^{n} (\tilde{s}_i^{max} + \tilde{s}_{i^*}^{max})\mathbb{E}[S_i(T)] + \tilde{s}_{i^*}^{max}, \\
&\leqslant \sum_{i=1, i\neq i^*}^{n} \Delta_i (C_1^i \log T + C_2^i) + \sum_{i=1, i\neq i^*}^{n} (\tilde{s}_i^{max} + \tilde{s}_{i^*}^{max}) \\
&\left(\log 2C_1^i \sqrt{\log_2 T} + (C_2^i + \log 2C_1^i)\left(1 + \frac{\pi^2}{6}\right) \right) + \tilde{s}_{i^*}^{max},
\end{aligned}$$

where

$$C_1^i = \frac{16\sigma_0^2}{\Delta_i^2} + \frac{\log 2}{2}\left(e^{\frac{3M_{i^*}^2}{2\sigma_0^2}} + e^{\frac{3M_i^2}{2\sigma_0^2}}\right),$$

$$C_2^i = \frac{4\sigma_0^2}{\Delta_i^2}\log\sqrt{2\pi e} + \left(e^{\frac{M_{i^*}^2}{3\sigma_0^2}} + e^{\frac{M_i^2}{3\sigma_0^2}}\right),$$

$\delta_i^2 = \sigma_0^2/\sigma_{i-cond}^2$, *and* $\sigma_{i-cond}^2 = \sigma_0^2 - \sigma_i(0)\Sigma_{\sim i}^{-1}(0)\sigma_i^T(0)$. $M_i = \sigma_0^2\sqrt{1+\delta_i^2}\sum_{j=1}^{n}\sum_{k=1}^{n}|\lambda_{kj}^0||\mu_j^0 - \mu_j|$ *measures the accuracy of the prior knowledge, where* λ_{kj}^0 *is the component of* Λ_0.

TABLE I
SIMULATION PARAMETERS

Parameters	**Value**
SBS transmit power	-10dBm ∼ 20dBm
MBS transmit power	40dBm
Thermal noise density	-174dBm/Hz
Bandwidth	20MHz
Carrier frequency	2GHz
Penetration loss(L_{ow})	20dB
Shadowing effect	log-normal with $\sigma = 8dB, \sigma' = 4dB$

$\tilde{s}_i^{max} = \max_{j=1,\dots,n} \mathbb{E}[s_{ij}]$ *is the maximum expected switching loss when SBS change power to* p_i.

The proof is omitted due to place limitations.

IV. SIMULATION RESULTS

We resort to numerical simulations to verify the effectiveness of the power assignment algorithm. A system-level simulator is developed, in which the indoor femto and outdoor macro channel model of urban deployment from [18] is used for UE-to-SBS and UE-to-MBS channels, respectively. Other important simulation parameters are summarized in Table I.

For this simulation setting, the SINR can be calculated as:

$$\mathsf{SINR}_{\mathsf{k_S,n}} = \frac{P^r_{k_S,n}}{\sum_{i=1,i\neq k_S}^{K} P^r_{i,n} + \sum_{j=1}^{K_M} P^{Mr}_{j,n} + N_s},$$

$$\mathsf{SINR}_{\mathsf{k_M,n}} = \frac{P^{Mr}_{k_M,n}}{\sum_{i=1}^{K} P^r_{i,n} \sum_{j=1,j\neq k_M}^{K_M} P^{Mr}_{j,n} + N_s},$$

for the inside and outside measurement points, respectively. $P^r_{i,n}$ and $P^{Mr}_{j,n}$ represent the received power at point n from SBS i and MBS j respectively. The PIF r under each power value can be calculated following the procedure in Sec. II.

In the single-SBS simulations, we simulate a warehouse with size 30×30 square meters. We set the center of the warehouse as origin and deploy a SBS at the grid point of $[12m, 8m]$ and an outside MBS at $[100m, 100m]$. The inside and outside routes follow the concentric circles pattern, whose radiuses are (2, 13) meters for the two indoor routes, and (24, 30) meters for the two outdoor routes. We set the time horizon as $T = 3000$ slots and run the system simulation to generate PIFs measured by r for each power value in $\mathcal{N}_{pow}$ with $\alpha = 0.7$. Here we adopt a simple linear function of switching loss as $s_{ij} = \gamma|p_i - p_j|$, where γ is a tunable parameter for different scenarios. Here we use the optimal power with maximum expected PIF achieved by the global optimization with complete RF information as the genie-aided optimum to evaluate the algorithm's performance.

We first verify the performance of CBPA-SC algorithm with different quality of prior knowledge. Fig. 3 reports the cumulative loss over time for CBPA-SC and two other algorithms – BPA-SC which utilizes only the prior estimations with a diagonal correlation matrix, and UiPA-SC which uses no prior knowledge. We can see that all three algorithms converge to the optimal power value asymptotically, but with different speed. Leveraging both the prior knowledge and the correlation structure significantly accelerates the convergence. Furthermore, in terms of minimizing the total PIF loss, CBPA-SC outperforms BPA-SC which performs better than UiPA-SC.

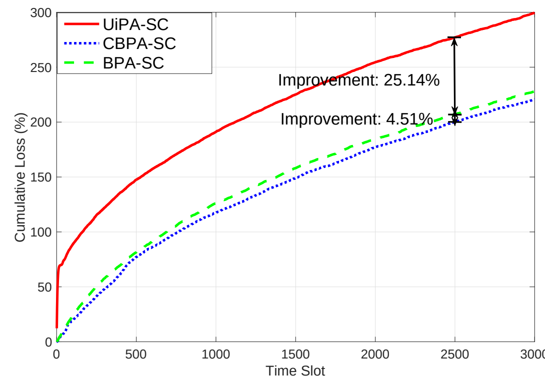

Fig. 3. Cumulative loss in a single-SBS deployment with $\alpha = 0.7$ and $\gamma = 0.2$.

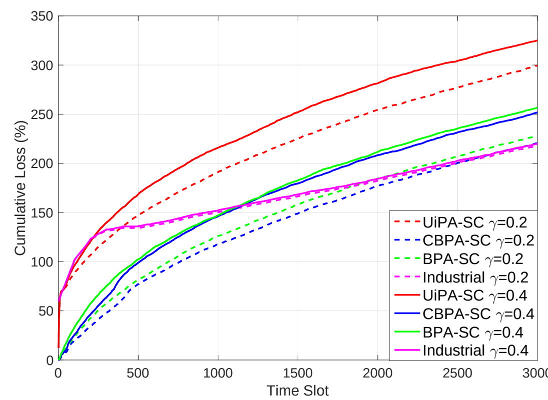

Fig. 4. Comparison of the industrial solution to UiPA-SC, BPA-SC and CBPA-SC in a large warehouse.

Next, we compare the proposed algorithm with the industry solution and report the results. The industrial heuristic sticks to a power value long enough to obtain the near-perfect PIF estimation, based on which it either increases or decreases the power value by some step size. Clearly, this method trades off fast convergence for certainty and is hardly influenced by switching cost. Fig. 4 and 5 reports the numerical comparison with a maximum 20dBm and step size 2dB. We can see that the industrial solution adapts poorly to different deployments, while our algorithms are stable due to the online learning and tradeoff between exploration and exploitation.

For the multi-SBS case, the size of the enterprise is set to be larger. The sample routes can be either ellipse or circle. We consider $K = 4$ in the simulations. The power value difference threshold is set to be $P_{th} = 5$dB. The power

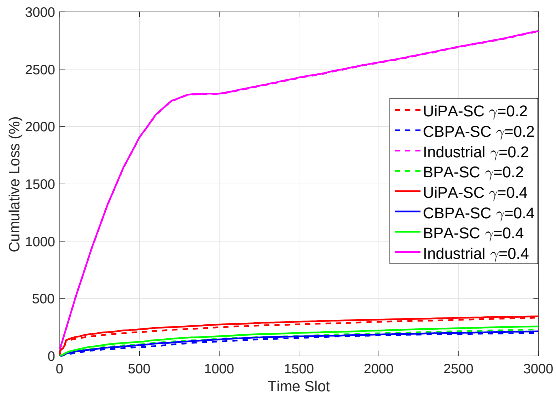

Fig. 5. Comparison of the industrial solution to UiPA-SC, BPA-SC and CBPA-SC in a small single-office enterprise.

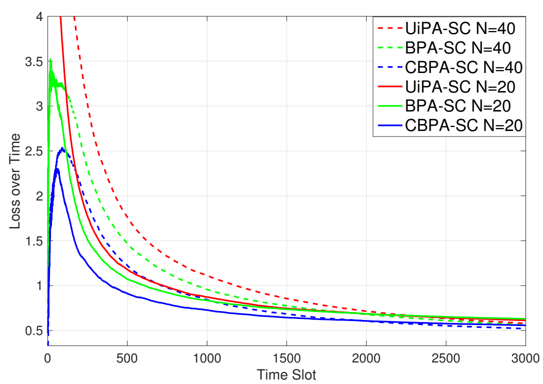

Fig. 6. Per-slot loss in $K = 4$ SBSs deployment with $\alpha = 0.7$ and $\gamma = 0.2$.

value for each SBS is selected from $\{-10, -5, .., 15, 20\}$dBm. It results in $n = 149$ power settings. We thus employ the clustering strategy in Sec. III-C and study two cases where the number of clusters is $N = 20$ and $N = 40$, respectively. The PIF loss normalized by time is shown in Fig. 6. We can see that all algorithms exhibit a decaying loss per slot. The effect of N can be analyzed from the figure. For all the three algorithms, larger cluster number results in worse performance in the initial period. This is because initially more power settings lead to more exploration and thus sub-optimal power settings are selected more. While a larger cluster number means one of the selected clustering medoids is closer to the global optimal power setting, a large N results in a better performance asymptotically.

V. Conclusion

We have studied the pilot power assignment problem associated with indoor enterprise closed-access SBS networks, in which the focus is on achieving optimal balance between providing sufficient coverage for the intended indoor users and suppressing leakage that causes interference to outdoor MBS users. We explicitly took into account the power switching cost, and proposed a block allocation scheme to reduce frequent power-switchings required for exploration. We formulated the performance criterion and modeled power assignment as an online learning problem. We adopted a Bayesian approach that leverages the prior information regarding the Gaussian distribution and proposed bandit-inspired power assignment algorithm. The CBPA-SC algorithm makes use of both prior knowledge of the mean and variance of each arm as well as the dependency of PIFs across different power values, outperforming other algorithms use less statistical information. Furthermore, a sub-linear upper bound for performance loss is proved for the algorithm. For the multi-SBS deployment, we proposed to use K-medoids clustering to reduce the complexity of the algorithm while maintaining the performance. When the cluster number is not very small, the algorithms can be close to the global optimal power setting for all K SBSs.

References

[1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020," February 2016.

[2] T. Quek, G. de la Roche, I. Guvenc, and M. Kountouris, *Small Cell Networks: Deployment, PHY Techniques, and Resource Allocation*. Cambridge University Press, 2013.

[3] O. Aliu, A. Imran, M. Imran, and B. Evans, "A survey of self organisation in future cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 336–361, 2013.

[4] Small Cell Forum, "Interference management in UMTS femtocells: topic brief," February 2014.

[5] S. Nagaraja *et al.*, "Downlink transmit power calibration for enterprise femtocells," in *IEEE VTC*, 2011.

[6] J. Kim, P.-Y. Kong, N.-O. Song, J.-K. K. Rhee, and S. Al-Araji, "MDP based dynamic base station management for power conservation in self-organizing networks," in *IEEE WCNC*, pp. 2384–2389, April 2014.

[7] M. Bennis *et al.*, "Self-Organization in Small Cell Networks: A Reinforcement Learning Approach," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 3202–3212, July 2013.

[8] S. Bubeck and N. Cesa-Bianchi, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Foundations and Trends in Machine Learning*, vol. 5, no. 1, pp. 1–122, 2012.

[9] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, pp. 235–256, May 2002.

[10] J. Zhu, Y. Song, D. Jiang, and H. Song, "Multi-armed bandit channel access scheme with cognitive radio technology in wireless sensor networks for the internet of things," *IEEE Access*, vol. 4, pp. 4609–4617, 2016.

[11] E. Kaufmann, O. Cappé, and A. Garivier, "On Bayesian upper confidence bounds for bandit problems," in *International Conference on Artificial Intelligence and Statistics*, pp. 592–600, April 2012.

[12] O. Atan, C. Tekin, and M. van der Schaar, "Global multi-armed bandits with Holder continuity," in *Proceedings of AISTATS*, 2015.

[13] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[14] T. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in Applied Mathematics*, vol. 6, pp. 4–22, 1985.

[15] V. Srivastava, P. Reverdy, and N. Leonard, "Correlated multiarmed bandit problem: Bayesian algorithms and regret analysis," *ArXiv e-prints*, July 2015, 1507.01160v2 [math.OC].

[16] R. Agrawal, M. V. Hegde, and D. Teneketzis, "Asymptotically efficient adaptive allocation rules for the multiarmed bandit problem with switching cost," *IEEE Trans. Autom. Control*, vol. 33, pp. 899–906, October 1988.

[17] H.-S. Park and C.-H. Jun, "A simple and fast algorithm for K-medoids clustering," *Expert Systems with Applications*, vol. 36, pp. 3336–3341, October 2009.

[18] 3GPP, "Evolved Universal Terrestrial Radio Access; Further advancements for E-UTRA physical layer aspects," TR 36.814.