Femtocell Scheduling as a Restless Multiarmed Bandit Problem Using Partial Channel State Observation

Hesham M. Elmaghraby*, Keqin Liu† and Zhi Ding*
*Dept. of Electrical and Computer Engineering, University of California, Davis, California, 95616
†KLA-Tencor, Milpitas, California, 95035

***Abstract*—[1] In this paper, we address the problem of channel allocation for femtocells that share the use of regular macrocell spectrum. The femto basestation (FBS) scheduling problem is formulated in the form of restless multiarmed bandit (RMAB) framework. Our goal is to choose the arms/channels that maximize the total expected discounted reward over infinite horizon while minimizing the induced interference due to channel sharing with macrocell. Without direct observation of true channel state, we use the available macrocell user feedback known as channel quality indicator (CQI). In general, the RMAB problem is P-SPACE hard. We propose a heuristic low complexity indexing policy referred as approximated Whittle index to rank available channels for FBS. Although finding a closed form channel ranking solution typically involve dynamic programming, we show that based on the partial channel information within CQI, there exists a closed form for the channel index. Moreover, we demonstrate the performance advantage of the proposed indexing policy over a myopic policy.**

***Index Terms*—Femtocell; restless multiarmed bandit (RMAB); resource allocation; Whittle index, myopic policy.**

I. Introduction

Multiarmed bandit (MAB) is a classical mathematical problem that provides basic framework for dynamic resource allocation problems. In its classical form, a player needs to choose one arm out of N to play and accordingly gains its reward. Each arm has certain state that determines its reward such that state transition can only occur for the active (chosen) arm. The player objective is to maximize its reward over infinite horizon through following a certain arm selection policy. Originally, MAB was first introduced in [1], and remained partially open since then. In [2] and [3], Gittins provided an essential insight to resolve the dimensionality problem in MAB by reducing the complexity from N-dimension problem into N-problems with 1-dimension for each.

A generalized form of the classical MAB formulation called the restless multiarmed bandit (RMAB) problem was studied by Whittle in [4]. In RMAB, the selection of K arms out of the available N arms is allowed ($1 \leq K \leq N$) instead of the previous one arm selection constraint, moreover the passive arms are allowed to change their state unlike the classical formulation where the state of any passive arm was unchanged. Whittle applied the Lagrangian relaxation to derive an indexing policy generalizing Gittins policy but for a much broader type of problems. Whittle index policy attached an index to each arm (referred as Whittle index), such that the player will choose the arms with the highest indices. The optimality of Whittle index policy was achieved by the nature of Lagrangian relaxation under a relaxed constraint on the mean number of chosen arms ($E[K(t)] = K$), instead of strict constraint $K(t) = K$.

Many real life applications can be viewed and modeled using the RMAB formulation. Despite the vast categories and types of problems, each problem has its own requirements and special considerations. Many existing works have linked the general channel allocation problem to the classical RMAB formulation. On the other hand, relatively fewer have applied this formulation on heterogeneous networks (HetNets) [5]–[8]. In this work, we employ the RMAB framework to model the femto basestation (FBS) resource allocation problem in HetNets where we take into account the existing feedback information. Our main motivation is to develop a practical and low complexity solution for the femtocell scheduling problem. In order to do so, we will need to provide practical considerations beyond the literature through using the RMAB formulation in scheduling the femtocell resources.

Femtocell resource scheduling is a typical problem that has been studied in many existing works [9]–[12]. It has been previously shown that the scheduling problem can be considered as a RMAB problem [5]–[7]. The ability to directly observe the channel state is one of the common assumptions made before which no longer applies in HetNet without direct and continuous coordination between macro basestation (MBS) and FBSs. Such coordination will require additional processing and will consume more bandwidth [6]. Unlike previous works, we did not assume that FBS can access the instantaneous channel state information. Instead, the proposed method utilizes the practical channel quality indicator (CQI) report as partial observation of the channel state. In HetNets, using cognitive capabilities, FBSs can observe the CQI based on which the channel state is estimated [13].

In this work, we aim to formulate the femtocell channel allocation problem in heterogeneous networks as a RMAB problem using practically available feedback information. We propose a low complexity solution by using an index policy for the FBS scheduling problem in which we utilize the overheard

[1]This material is based on work supported by the National Science Foundation under Grants CNS 1443870 and CNS 1702752. The work of the 1^{st} author is also supported by an Egyptian Government grant.

978-1-5386-3180-5/18/$31.00 ©2018 IEEE

CQI feedback information. Further, we derive a closed form index for the given problem using an approximated belief value. Our performance gain is shown by comparing the proposed policy, and myopic policy.

Our manuscript is organized as follows, Section II introduces the system model and assumptions used throughout this work. In Section III, we present the RMAB problem formulation based on the received CQI observation. We provide our index derivation as well as its closed form in Section IV. Section V provides comparisons between the performance of the proposed indexing policy versus a myopic policy. Lastly, we conclude our work in Section VI.

II. System Model

Our network model consists of a central MBS, owned by the service provider, and number of FBSs processed by the femto holders as shown in Fig. 1. Each FBS shares a number of assigned channels with the MBS such that the co-tier interference with other FBSs is avoided. This can be achieved by orthogonal channel assignments for adjacent femtocells. Further, we assume cognitive capabilities in each FBS in order to assist in overhearing the CQI report. Furthermore, femtocells operate in closed access mode and the physical layer follows the LTE-A time division duplex (TDD) frame structure. Owing to channel reciprocity, the channel gains between the FBS and the users (femto or macro users) are considered to be known, as well as the channel connecting the MBS and the femto users (FUEs). On the other hand, the channel connecting the MBS and the macro users (MUEs) is considered as unknown channel for the FBS. Moreover, we require no periodic direct information transfer between FBS and MBS to finish the scheduling process.

We assume that the channel model describing unknown channels follow the famous 2-state Markov model (Gilbert-Elliott model) shown in Fig. 1. Each channel has its transition probabilities P_{bg}, P_{gg} and the probability of being in the good/bad state P_g, P_b.

According to our system model, each FBS is assigned a group of orthogonal channels, such that FBSs can obtain their observation through overhearing the CQI report ($q(t)$) for the assigned channels. The observed CQI report includes the MUEs signal to interference and noise ratio (SINR) information. However, full observation is obtained by monitoring the CQI for the chosen subset of channels, in order to measure the impact of channel sharing on the primary users (MUEs) received signal quality such that the instantaneous observed CQI for a certain channel at time t is denoted as $q(t)$. The number of available CQI levels will be referred as N_q such that $1 \leq q(t) \leq N_q$.

III. Problem Formulation

We will provide a generalization for regular RMAB formulation presented in [6], such that instead of directly observing the real channel state, we will use the available feedback information (CQI) embedded in the users channel state information

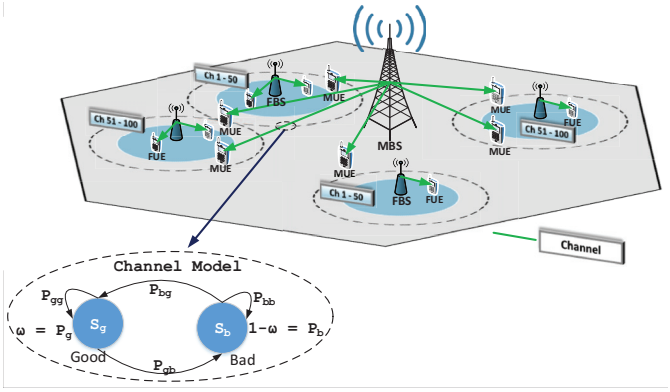

Fig. 1. System model.

(CSI). We will assume that the FBS has N_c independent channels shared with neighboring MUEs.

Each channel is modeled using the Gilbert-Elliott channel model shown in Fig. 1. At each time, FBS has to schedule K channels for K FUEs. We aim to use Whittle index to rank the available channels based on the subsidy for passivity concept [4], such that the more attractive the channel for use the higher rank or index it will have.

The belief $\omega(t)$ refers to the conditional probability of being in the good state at time t. While the belief vector ($\boldsymbol{\omega}(t)$) consists of all the instantaneous beliefs of the available channels. For channel i, denote $a_i(t)$ as the decided action based on the observed CQI according to certain policy function, while $\boldsymbol{a}(t)$ is the action vector for all the available channels. Basically, the policy function (π) will map the belief vector to the action.

Our objective is to maximize the expected discounted reward over infinite horizon which can be described by

$$\mathrm{E}_{\pi}\left(\sum_{t=1}^{\infty}\delta^{t-1}R_{\pi}(t)|\boldsymbol{\omega}(1)\right), \tag{1}$$

where $\boldsymbol{\omega}(1)$ is the initial belief vector, δ is the discount factor ($\delta \in [0\ 1]$), and the reward function is defined as

$$R_{\pi}(t) = \sum_{i \in A(t)} X_i(t) B_i, \tag{2}$$

where $A(t)$ is the set of active arms/channels at time t, $X_i(t)$ is state indicator function equals to "1" when the channel is in the good state (S_g) and "0" otherwise. B_i is the throughput/bandwidth of channel i. $\boldsymbol{B} = \{B_1, B_2 \ldots B_N\}$ is the vector containing the channels throughputs, and without loss of generality, we normalized this vector such that the greatest throughput is 1. In the next section, we will drop the channel index, since the proposed indexing policy will handle each channel independently (i.e. $B_i = 1$).

A. Restless Multiarmed Bandit (RMAB) Problem Statement

According to our system model, each FBS is assigned number of channels to schedule their users, such that for FBS F there exists N_c channels and the FBS need to assign K channels for K FUEs. We assume that the MBS-MUEs

channels follow Gilbert-Elliott channel model shown in Fig. 1, meanwhile these channels are considered as unknown channels to the FBS, hence their actual state is unobservable. Instead of observing the channels actual state, FBSs can overhear/observe the CQI reports for the available channels, in order to update the probability of being in the good state (belief). Accordingly, we need to update the belief based on the available observation (i.e. $q(t)$) as follows

$$\omega(t+1) = \begin{cases} \omega_i, \text{ for } a(t)=1, q(t)=i \\ T[\omega(t)], \text{ for } a(t)=0 \end{cases}, \quad (3)$$

where ω_i is the probability that the channel state at $t+1$ is in the good state when the observation ($q(t)$) at time t is i ($\Pr(S(t+1)=S_g|q(t)=i)$) for $i=1,\dots N_q$ and $T[\omega(t)]$ is the belief update when passive action is decided, hence no observation such that

$$T[\omega(t)] = p_{gg}\omega(t) + p_{bg}(1-\omega(t)). \quad (4)$$

Define the belief of receiving a CQI observation i as

$$\omega_i = \Pr(S(t+1)=S_g|q(t)=i), \quad (5)$$

and using the law of total probability, we get

$$\begin{aligned} \omega_i =& \Pr(S(t+1)=S_g|S(t)=S_g)\Pr(S(t)=S_g|q(t)=i) \\ &+\Pr(S(t+1)=S_g|S(t)=S_b)\Pr(S(t)=S_b|q(t)=i), \\ \omega_i =& \frac{p_{gg}p_{ig}\omega(t) + p_{bg}p_{ib}(1-\omega(t))}{p_{ig}\omega(t) + p_{ib}(1-\omega(t))} \quad \text{for } i=1\dots N_q, \end{aligned} \quad (6)$$

while the probability of observing a certain CQI given $\omega(t)$ ($\Pr(q(t)=i|\omega(t))$) is given by

$$p_i(\omega) = \omega p_{ig} + (1-\omega)p_{ib}, \quad (7)$$

where p_{ig}, p_{ib} are the probabilities of observing $q(t)=i$ when the channel is in the good or bad state, respectively. Accordingly $\sum_{i=1}^{N_q} p_{ig} = 1$ and $\sum_{i=1}^{N_q} p_{ib} = 1$.

The value function ($f_{s,\delta}(\omega)$) represents the maximum expected total discounted reward achievable using a single armed bandit with a subsidy s and belief ω. It can be described as

$$f_{s,\delta}(\omega) = \max(f_{s,\delta}(\omega;a=0), f_{s,\delta}(\omega;a=1)), \quad (8)$$

where $f_{s,\delta}(\omega;a=0)$ is the value function when a passive action is taken (no observation) while $f_{s,\delta}(\omega;a=1)$ is the value function when an active action is taken and the discount factor is δ. We denote the value function for the passive and active actions as

$$f_{s,\delta}(\omega;a=0) = s + \delta f_{s,\delta}(T[\omega]), \quad (9)$$

$$f_{s,\delta}(\omega;a=1) = \omega + \delta(\sum_{i=1}^{N_q} p_i(\omega) f_{s,\delta}(\omega_i)), \quad (10)$$

where s is the subsidy that represent the reward for the passive action in the first time slot.

The RMAB is considered indexable if all the arms/channels are indexable. Due to the problem decomposability, showing that an arbitrary arm is indexable is suffice to prove that all arms are indexable, hence the problem is indexable. According to [4] and [6], an arm is considered indexable if the passive set of a single armed bandit process with a certain subsidy s increases from $\varnothing$ to include the whole space $[0,1]$ monotonically with the increase of s from $-\infty$ to ∞.

For indexable problem, the Whittle index can be defined as

$$W(\omega) = \inf_s \{s : f_{s,\delta}(\omega;a=0) = f_{s,\delta}(\omega;a=1)\}, \quad (11)$$

where $W(\omega)$ is Whittle index at belief ω. In this work, we did not prove the problem indexability, so our indexing policy will be referred as approximated Whittle index policy and our approximated index ($\bar{W}(\omega)$) will be defined as in (11)

$$\bar{W}(\omega) = \inf_s \{s : f_{s,\delta}(\omega;a=0) = f_{s,\delta}(\omega;a=1)\}. \quad (12)$$

We will base our solution on a threshold policy such that starting from any belief value, arms will reach certain threshold which it will be easier to define the value function. Such policy was previously presented in [6]. First, we need to show that there exists a certain threshold $w^*(s)$ where the passive and active actions are equally attractive.

We need to emphasis that the belief propagation for the unobserved arm can be derived using the same procedure shown in [6], such that

$$T^k[\omega] = \frac{p_{bg} - (p_{gg}-p_{bg})^k(p_{bg} - (1+p_{bg}-p_{gg})\omega)}{1+p_{bg}-p_{gg}}, \quad (13)$$

where k is the number of the propagated time slots. Another important function to define is $L(\omega,\nu)$, which represents the time taken for a passive arm to propagate from a certain belief (ω) to another one (ν). Since this function evaluates the time for a passive arm, then the belief update will follow the passive arm update shown in (3), leading to the result introduced in [6], at $p_{gg} \geq p_{bg}$

$$L(\omega,\nu) = \begin{cases} 0, & \text{if } \omega > \nu \\ \lfloor \log_{p_d} \frac{p_{bg}-\nu(1-p_d)}{p_{01}-\omega(1-p_d)} \rfloor + 1, & \text{if } \omega \leq \nu < \omega_{ss} \\ \infty, & \text{if } \omega \leq \nu \ \& \ \nu \geq \omega_{ss} \end{cases}, \quad (14)$$

where $p_d = p_{gg} - p_{bg}$ and $\omega_{ss} = p_{bg}/(1+p_{bg}-p_{gg})$ represents steady state belief, and for $p_{bg} > p_{gg}$

$$L(\omega,\nu) = \begin{cases} 0, & \text{if } \omega > \nu \\ 1, & \text{if } \omega \leq \nu \text{ and } T[\omega] > \nu \\ \infty, & \text{if } \omega \leq \nu \text{ and } T[\omega] \leq \nu \end{cases}. \quad (15)$$

In the next section, we need to construct our policy structure.

B. Policy Structure

In this section, we need to show that there exists a threshold $\omega^*(s)$ such that the arm will transfer from the passive state to the active state where $f_{s,\delta}(\omega;a=0) = f_{s,\delta}(\omega;a=1)$.

We will show that for $0 \leq s < 1$, there must exists at least one belief value such that the passive action value function is equal to the value function when the arm is active. From (6), we can deduce that $\omega_i(t)|_{\omega(t)=0} = p_{bg}$ and $\omega_i(t)|_{\omega(t)=1} = p_{gg}$, while from (7), we have $p_i(\omega)|_{\omega(t)=0} = p_{ib}$ and $p_i(\omega)|_{\omega(t)=1} = p_{ig}$ for $i=1\dots N_q$.

Then for $0 \leq s < 1$:
At $\omega(t) = 0$

$$\begin{aligned} f_{s,\delta}(\omega = 0; a = 1) =& \delta \sum_{i=1}^{N_q} p_{ib} f_{s,\delta}(p_{bg}) = \delta f_{s,\delta}(p_{bg}) \sum_{i=1}^{N_q} p_{ib} \\ f_{s,\delta}(\omega = 0; a = 1) =& \delta f_{s,\delta}(p_{bg}) \leq s + \delta f_{s,\delta}(p_{bg}) \\ & = f_{s,\delta}(\omega = 0; a = 0) \\ f_{s,\delta}(\omega = 0; a = 1) \leq & f_{s,\delta}(\omega = 0; a = 0). \end{aligned} \tag{16}$$

While at $\omega(t) = 1$

$$\begin{aligned} f_{s,\delta}(\omega = 1; a = 1) =& 1 + \delta \sum_{i=1}^{N_q} p_{ig} f_{s,\delta}(p_{gg}) \\ f_{s,\delta}(\omega = 1; a = 1) =& 1 + \delta f_{s,\delta}(p_{gg}) > s + \delta f_{s,\delta}(p_{gg}) \\ & = f_{s,\delta}(\omega = 1; a = 0) \\ f_{s,\delta}(\omega = 1; a = 1) > & f_{s,\delta}(\omega = 1; a = 0). \end{aligned} \tag{17}$$

From which we can confirm that for $0 \leq s < 1$ there exists at least one crossing such that $f_{s,\delta}(\omega; a = 0) = f_{s,\delta}(\omega; a = 1)$ at some point $\omega^*(s)$.

IV. Closed Form Value Function

In order to derive a close form for the value function, we need to refer to a base case for which we reach a certain constant belief. According to our optimum policy and the value function structure, we can describe the value function at any belief to start with the value function for the passive action until we reach $\omega^*(s)$ then the arm will be transfered from the passive to active action.

Accordingly, we can write the value function in terms of the active action value function after $L(\omega, \omega^*(s)) + 1$ time slots. $L(\omega, \omega^*(s))$ is the time needed to transfer from the starting belief (ω) until it reaches $\omega^*(s)$, such that

$$\begin{aligned} f_{s,\delta}(\omega) =& \frac{1 - \delta^{L(\omega,\omega^*(s))}}{1 - \delta} s \\ & + \delta^{L(\omega,\omega^*(s))} f_{s,\delta}(T^{L(\omega,\omega^*(s))}[\omega]; a = 1), \end{aligned} \tag{18}$$

which have the same structure as the value function shown in [6], as outcome of the threshold policy similarity. Since $f_{s,\delta}(T^{L(\omega,\omega^*(s))}(\omega); a = 1)$ is function in ω_i, then we can evaluate a close form for $f_{s,\delta}(\omega_i)$ which can be used to evaluate a general closed form for $f_{s,\delta}(\omega)$. According to (6), ω_i is function in $\omega(t)$ which prevent having the base case needed to derive a closed form.

In order to recover the problem of having ω_i as function in $\omega(t)$, we are suggesting an approximation where instead of using the instantaneous belief value ($\omega(t)$), we use the steady state belief value (ω_{ss}) such that

$$\omega_i(t) = \frac{p_{gg} p_{ig} \omega(t) + p_{bg} p_{ib} (1 - \omega(t))}{p_{ig} \omega(t) + p_{ib} (1 - \omega(t))}, \tag{19}$$

can be approximated to

$$\omega_i(t) \approx \ \bar{\omega}_i = \frac{p_{gg} p_{ig} \omega_{ss} + p_{bg} p_{ib} (1 - \omega_{ss})}{p_{ig} \omega_{ss} + p_{ib} (1 - \omega_{ss})}, \tag{20}$$

where $\omega_{ss} = p_{bg}/(1 + p_{bg} - p_{gg})$. By this way we can reach a constant ω_i and have the base case needed to derive a closed form value function.

From (19), we can see that the values of $\omega_i(t)$ varies between p_{bg} to p_{gg} such that $\omega_i(t)|_{\omega(t)=0} = p_{bg}$ and $\omega_i(t)|_{\omega(t)=1} = p_{gg}$.

We will start by substituting in (18) to evaluate $f_{s,\delta}(\bar{\omega}_i)$ to get

$$f_{s,\delta}(\bar{\omega}_i) = x_i s + y_i + \sum_{j=1, j \neq i}^{N_q} z_{i,j} f_{s,\delta}(\bar{\omega}_j), \tag{21}$$

where

$$\begin{aligned} x_i =& \frac{1 - \delta^{L(\bar{\omega}_i,\omega^*(s))}}{(1 - \delta^{L(\bar{\omega}_i,\omega^*(s))+1} p_i(T^{L(\bar{\omega}_i,\omega^*(s))}[\bar{\omega}_i]))(1 - \delta)} \\ y_i =& \frac{\delta^{L(\bar{\omega}_i,\omega^*(s))} T^{L(\bar{\omega}_i,\omega^*(s))}[\bar{\omega}_i]}{1 - \delta^{L(\bar{\omega}_i,\omega^*(s))+1} p_i(T^{L(\bar{\omega}_i,\omega^*(s))}[\bar{\omega}_i])}, \\ z_{i,j} =& \frac{\delta^{L(\bar{\omega}_i,\omega^*(s))+1} p_j(T^{L(\bar{\omega}_i,\omega^*(s))}[\bar{\omega}_i])}{1 - \delta^{L(\bar{\omega}_i,\omega^*(s))+1} p_i(T^{L(\bar{\omega}_i,\omega^*(s))}[\bar{\omega}_i])}, \end{aligned}$$

now we got N_q equations for $f_{s,\delta}(\bar{\omega}_i)$ for $i = 1 \ldots N_q$, which can be written in matrix form as a system of linear equations

$$\begin{bmatrix} 1 & -z_{1,2} & \cdots & -z_{1,N_q} \\ -z_{2,1} & 1 & \cdots & -z_{2,N_q} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ -z_{N_q,1} & -z_{N_q,2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} f_{s,\delta}(\bar{\omega}_1) \\ f_{s,\delta}(\bar{\omega}_2) \\ \vdots \\ \vdots \\ f_{s,\delta}(\bar{\omega}_{N_q}) \end{bmatrix} = \begin{bmatrix} x_1 s + y_1 \\ x_2 s + y_2 \\ \vdots \\ \vdots \\ x_{N_q} s + y_{N_q} \end{bmatrix}, \tag{22}$$

then

$$\bar{A} \bar{\boldsymbol{v}} = \bar{\boldsymbol{b}}, \tag{23}$$

where $\bar{A}, \bar{\boldsymbol{v}}$ and $\bar{\boldsymbol{b}}$ are the matrix and arrays shown in (22).

This set of linear equations can be solved in polynomial time ($\mathcal{O}(n^3)$), but the value functions will still be function in the subsidy s. The number of equations in (22) are N_q, meanwhile we got $N_q + 1$ unknown (N_q value functions and the subsidy s). Accordingly, we need to add one more equation to the equation set in (22) to be uniquely solvable. Using the equality condition of the active and passive value functions, we should be able to add the needed equation, as follows

$$f_{s,\delta}(\omega; a = 0) = f_{s,\delta}(\omega; a = 1), \tag{24}$$

$$s + \delta f_{s,\delta}(T[\omega]) = \omega + \delta \sum_{i=1}^{N_q} p_i(\omega) f_{s,\delta}(\bar{\omega}_i), \tag{25}$$

and substituting with the results illustrated in (18), we will get

$$s - \sum_{i=1}^{N_q} \epsilon_i f_{s,\delta}(\bar{\omega}_i) = \eta, \tag{26}$$

where

$$\begin{aligned} \eta =& \frac{(1 - \delta)(\omega - \delta^{L(T[\omega],\omega^*(s))+1} T^{L(T[\omega],\omega^*(s))}[T[\omega]])}{1 - \delta^{L(T[\omega],\omega^*(s))+1}}, \\ \epsilon_i =& \frac{\delta(1 - \delta) p_i(\omega) - \delta^{L(T[\omega],\omega^*(s))+1} p_i(T^{L(T[\omega],\omega^*(s))}[T[\omega]])}{1 - \delta^{L(T[\omega],\omega^*(s))+1}}. \end{aligned}$$

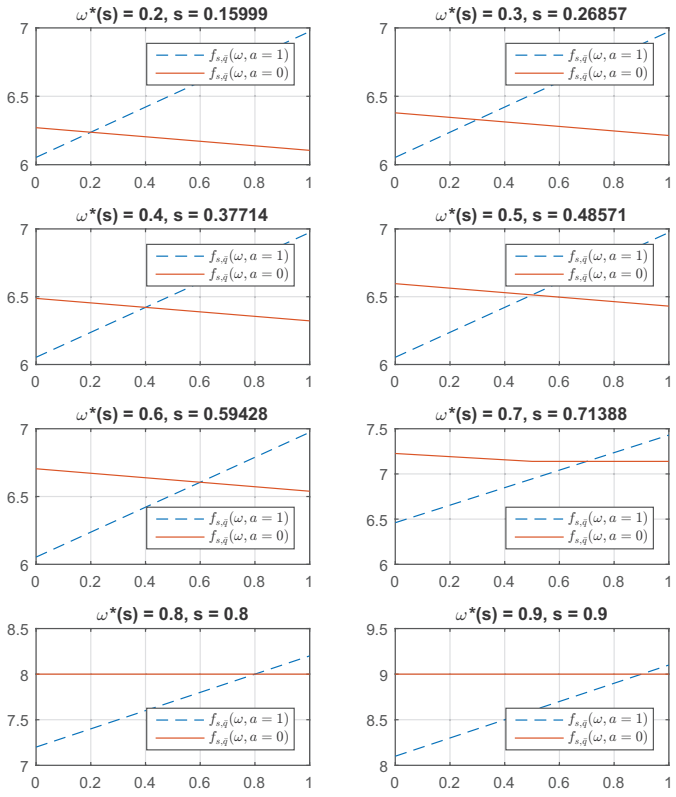

Fig. 2. Optimum threshold for different $\omega^*(s)$.

Using (26), we can evaluate Whittle index directly from the given equations as follows

$$\begin{bmatrix} & \bar{A} & & -x_1 \\ & & & \vdots \\ & & & -x_{N_q} \\ -\epsilon_1 & \cdots & -\epsilon_{N_q} & 1 \end{bmatrix} \begin{bmatrix} f_{s,\delta}(\bar{\omega}_1) \\ f_{s,\delta}(\bar{\omega}_2) \\ \vdots \\ f_{s,\delta}(\bar{\omega}_{N_q}) \\ s \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N_q} \\ \eta \end{bmatrix}, \tag{27}$$

such that

$$A\boldsymbol{v} = \boldsymbol{b}, \tag{28}$$

where $A, \boldsymbol{v}$ and $\boldsymbol{b}$ are the matrix and arrays shown in (27). Now, we can calculate the approximated Whittle index by solving the set of linear equations described in (27) and (28) with polynomial complexity ($\mathcal{O}(n^3)$), where n is the number of unknowns ($n = N_q + 1$), or by applying Cramer's rule with the same complexity Thus, the approximated Whittle index for each channel i with a bandwidth B_i can be described as follows

$$\bar{W}_i(\omega) = B_i \frac{\det(A_s)}{\det(A)}\bigg|_{\omega^*(s)=\omega}. \tag{29}$$

From (29), we can deduce that the existence of the approximated Whittle index is directly related to the matrix A non-singularity, that is to say that there exists an approximated Whittle index when the matrix A is full rank.

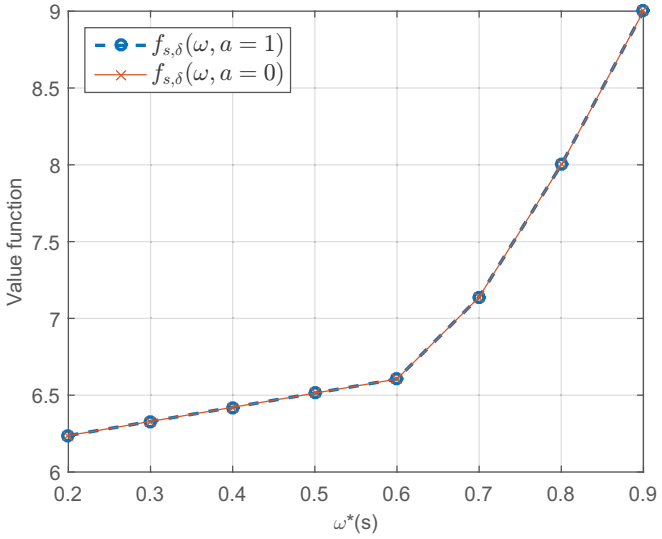

Fig. 3. Active and passive value functions at Whittle index.

V. Numerical Performance Results

We divide the performance section into two main parts, part I will focus on validating the presented formulas in the previous sections, while in part II, we will compare the performance of the proposed heuristic index policy and the myopic policy.

Figs. 2, and 3 demonstrates the existence of a Whittle index and correctness for different belief values. Fig. 4, will illustrate the performance of the approximated Whittle policy compared to the myopic policy.

In Fig. 2, we provide the value function at $a = 0$ versus $a = 1$, to show the existence of approximated Whittle index for different belief values. It is clear from the figures that the intersection point ($\omega^*(s)$) perfectly match the calculated value.

In Fig. 3, we present the value function variation with $\omega^*(s)$ at the approximated Whittle index which confirms the correctness of the calculated index such that it satisfy the active and passive value functions equality condition in (24).

A. Myopic Policy Comparison

The myopic policy is considered to be one of the simplest non-trivial RMAB polices, where the objective is to maximize the current reward only without considering future rewards. The myopic action ($\hat{A}(\boldsymbol{\omega})$) for a belief vector $\boldsymbol{\omega}$ is

$$\hat{A}(\boldsymbol{\omega}) = \arg\max_{A(t)} \sum_{i \in A(t)} \omega^i B_i, \tag{30}$$

where ω^i is the instantaneous belief of channel i. The simulation flow used to evaluate the system performance start by generating the channel states according to the provided channel model, then we generate the CQI observations in accordance. Based on the CQI observation, we calculated the actual belief using (6), which will be used by the myopic policy to select the used arm. In (20), we provide an approximation for the belief which was used to derive our Whittle policy.

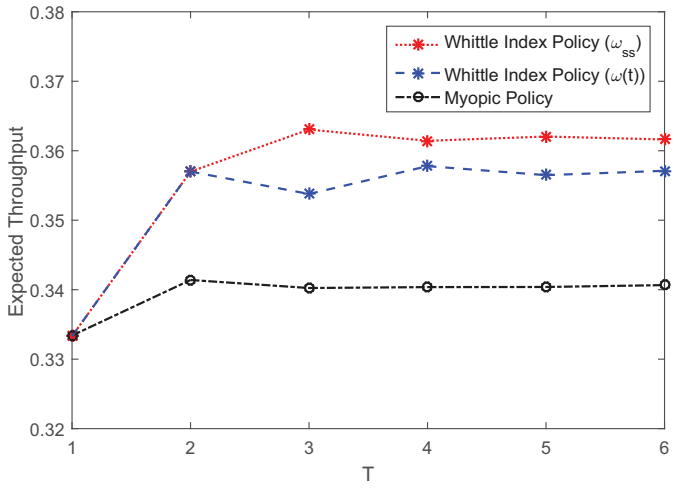

Fig. 4. Expected throughput using the Whittle and myopic policies $(\{p_{bg}^{(i)}\}_{i=1}^{7} = \{0.8, 0.6, 0.4, 0.9, 0.8, 0.6, 0.7\}$, $\{p_{gg}^{(i)}\}_{i=1}^{7} = \{0.6, 0.4, 0.2, 0.2, 0.4, 0.1, 0.3\}$, $N_q = 15, K = 1, N = 7$ and $B_i = \{0.4998, 0.6668, 1.0000, 0.6296, 0.5830, 0.8334, 0.6668\}$.

In Fig. 4, we present the expected throughput for the myopic and Whittle indexing policies. The dashed blue lines provide the expected throughput using the actual belief update, while the dotted red lines shows the expected throughput when using the approximated belief. Although, we used the actual belief in generating the myopic policy response, Whittle index policy which is based on the approximated belief still outperform the myopic policy. The expected reward for the myopic and Whittle indexing policies presented in Fig. 4 is calculated using $N_q = 15$ and the reward is defined as

$$R_{A(t)}(t) = \sum_{i \in A(t)} X_i(t) B_i, \tag{31}$$

such that if the state of the observed channel is good then $X_i(t) = 1$, and $X_i(t) = 0$ otherwise. In other words, the user transmits and collects reward if the selected channel is in the good state, otherwise, the user collects no reward. This performance measure is based on the channel state, which should be a common measure between the myopic and Whittle index policies. In Fig. 4, we evaluated the throughput for the proposed Whittle index policy using the approximated steady state belief and the actual belief, which shows partially the approximated steady state belief effect on the presented formula.

VI. Conclusion

In this work, we formulated the FBS channel allocation problem in heterogeneous networks where the femtocell available bandwidth is shared with the macrocell. We deployed the RMAB framework in order to describe the FBS scheduling problem while taking into account the available feedback information. We leverage the RMAB formulation to develop an indexing policy that represents a practical low complexity solution for the shared FBS resource allocation problem. We proved the existence of an optimal belief where the active and passive actions value functions are equally attractive under the new problem formulation. Moreover, we derived an approximate closed form Whittle index for the given problem using an approximate belief value. Lastly, we illustrated the advantage of the proposed Whittle index policy through test comparisons with results of the myopic policy.

References

[1] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 25, no. 3/4, pp. 285–294, 1933.

[2] D. M. Jones and J. C. Gittins, *A dynamic allocation index for the sequential design of experiments*. University of Cambridge, Department of Engineering, 1972.

[3] J. C. Gittins, "Bandit processes and dynamic allocation indices," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 148–177, 1979.

[4] P. Whittle, "Restless bandits: Activity allocation in a changing world," *Journal of applied probability*, pp. 287–298, 1988.

[5] K. Liu and Q. Zhao, "Distributed learning in cognitive radio networks: Multi-armed bandit with distributed multiple players," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 3010–3013.

[6] ——, "Indexability of restless bandit problems and optimality of whittle index for dynamic multichannel access," *Information Theory, IEEE Transactions on*, vol. 56, no. 11, pp. 5547–5567, 2010.

[7] Y. Gai and B. Krishnamachari, "Distributed stochastic online learning policies for opportunistic spectrum access," *Signal Processing, IEEE Transactions on*, vol. 62, no. 23, pp. 6184–6193, 2014.

[8] A. Feki and V. Capdevielle, "Autonomous resource allocation for dense lte networks: A multi armed bandit formulation," in *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2011, pp. 66–70.

[9] L. Zhang, L. Yang, and T. Yang, "Cognitive interference management for LTE-A femtocells with distributed carrier selection," in *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, Sept 2010, pp. 1–5.

[10] R. Xie, F. Yu, and H. Ji, "Spectrum sharing and resource allocation for energy-efficient heterogeneous cognitive radio networks with femtocells," in *Communications (ICC), 2012 IEEE International Conference on*, June 2012, pp. 1661–1665.

[11] S. Lien, Y. Lin, and K. Chen, "Cognitive and game-theoretical radio resource management for autonomous femtocells with QoS guarantees," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 7, pp. 2196–2206, July 2011.

[12] D. López-Pérez, X. Chu, A. V. Vasilakos, and H. Claussen, "Power minimization based resource allocation for interference mitigation in OFDMA femtocell networks," *Selected Areas in Communications, IEEE Journal on*, vol. 32, no. 2, pp. 333–344, 2014.

[13] H. M. Elmaghraby, D. Qin, and Z. Ding, "Downlink scheduling and power allocation in cognitive femtocell networks," in *Cognitive Radio Oriented Wireless Networks*. Springer, 2015, pp. 92–105.