Augmenting Max-Weight with Explicit Learning for Wireless Scheduling with Switching Costs

Subhashini Krishnasamy*, Akhil P T†, Ari Arapostathis*, Sanjay Shakkottai* and Rajesh Sundaresan‡†
*Department of ECE, The University of Texas at Austin
Email: subhashini.kb@utexas.edu, ari@ece.utexas.edu, shakkott@mail.utexas.edu
†Department of ECE, Indian Institute of Science
Email: akhilpt@gmail.com, rajeshs@ece.iisc.ernet.in
‡Robert Bosch Centre for Cyber Physical Systems, Indian Institute of Science

***Abstract*—In small-cell wireless networks where users are connected to multiple base stations (BSs), it is often advantageous to opportunistically switch off a subset of BSs to minimize energy costs. We consider two types of energy cost: (i) the cost of maintaining a BS in the active state, and (ii) the cost of switching a BS from the active state to inactive state. The problem is to operate the network at the lowest possible energy cost (sum of activation and switching costs) subject to queue stability. In this setting, the traditional approach – a Max-Weight algorithm along with a Lyapunov-based stability argument – does not suffice to show queue stability, essentially due to the temporal co-evolution between channel scheduling and the BS activation decisions induced by the switching cost. Instead, we develop a learning and BS activation algorithm with slow temporal dynamics, and a Max-Weight based channel scheduler that has fast temporal dynamics. We show using convergence of time-inhomogeneous Markov chains, that the co-evolving dynamics of learning, BS activation and queue lengths lead to near optimal average energy costs along with queue stability.**

***Index Terms*—wireless scheduling, base-station activation, energy minimization**

I. Introduction

Due to tremendous increase in demand for data traffic, modern cellular networks have taken the densification route to support peak traffic demand [1]. While increasing the density of base-stations gives greater spectral efficiency, it also results in increased costs of operating and maintaining the deployed base-stations. Rising energy cost is a cause for concern, not only from an environmental perspective, but also from an economic perspective for network operators as it constitutes a significant portion of the operational expenditure. To address this challenge, latest research aims to design energy efficient networks that balance the trade-off between spectral efficiency, energy efficiency and user QoS requirements [2], [3].

Studies reveal that base-stations contribute to more than half of the energy consumption in cellular networks [4], [5]. Although dense deployment of base-stations are useful in meeting demand in peak traffic hours, they regularly have excess capacity during off-peak hours [3], [6]. A fruitful way to conserve power is, therefore, to dynamically switch off under-utilized base-stations. For this purpose, modern cellular standards incorporate protocols that include *sleep* and *active* modes for base-stations. The sleep mode allows for selectively switching under-utilized base-stations to low energy consumption modes. This includes completely switching off base-stations or switching off only certain components.

Consider a time-slotted multi base-station (BS) cellular network where subsets of BSs can be dynamically activated. Since turning off BSs could adversely impact the performance perceived by users, it is important to consider the underlying energy vs. performance trade-off in designing BS activation policies. In this paper, we study the joint problem of dynamically selecting the BS activation sets and user rate allocation depending on the network load. We take into account two types of overheads involved in implementing different activation modes in the BSs.

(i) Activation cost occurs due to maintaining a BS in the active state. This includes energy spent on main power supply, air conditioning, transceivers and signal processing [6]. Surveys show that a dominant part of the energy consumption of an active base-station is due to static factors that do not have dependencies with traffic load intensities [3], [7]. Therefore, an active BS consumes almost the same energy irrespective of the amount of traffic it serves. Typically, the operation cost (including energy consumption) in the sleep state is much lower than that in the active state since it requires only minimal maintenance signaling [5].

(ii) Switching cost is the penalty due to switching a BS from active state to sleep state or vice-versa. This factors in the signaling overhead (control signaling to users, signaling over the backhaul to other BSs and/or the BS controller), state-migration processing, and switching energy consumption associated with dynamically changing the BS modes [6].

Further, switching between these states typically cannot occur instantaneously. Due to the hysteresis time involved in migrating between the active and sleep states, BS switching can be done only at a slower time-scale than that of channel scheduling [8], [9].

Main Contributions

We formulate the problem in a (stochastic) network cost minimization framework. The task is to select the set of active BSs in every time-slot, and then based on the instantaneous channel state for the activated BSs, choose a feasible allocation of rates to users. Our aim is to minimize the total network cost (sum of activation and switching costs) subject to stability of the user queues at the BSs.

Insufficiency of the standard Lyapunov technique: Such stochastic network resource allocation problems typically adopt greedy primal dual algorithms along with virtual-queues to accommodate resource constraints [10], [11], [12]. To ensure stability, this technique crucially relies on achieving negative Lyapunov drift in every time-slot (or within some finite number of time-slots). This is feasible in the traditional setting because the channel state in every time-slot is independent of the controller's actions, and therefore, provides an average (potential) service rate that is strictly higher than the average arrival rate.

In our problem, unlike in the traditional setting, one-step greedy Lyapunov based algorithms cannot be directly used. Recall that the cost at each time-slot has two components: *(a)* activation cost, and *(b)* switching cost. Further, effective channel state in each time-slot (consisting of feasible rates for the activated BS set) is determined by the BS activation decision in that time-slot. Since switching cost depends on the change in activation state in consecutive time-slots, traditional virtual-queue based algorithms introduce coupling of activation decisions across time. Thus, the evolution of the effective channel rates are dependent across time through the scheduling decisions, and this results in co-evolution of packet queues and the channel state distribution.

To circumvent difficulties introduced through this co-evolution, we propose an approach that uses queue-lengths for channel scheduling at a fast time-scale, but explicitly uses arrival and channel statistics (using learning via an explore-exploit learning policy) for activation set scheduling at a slower time-scale. Our main contributions are as follows.

1) **Static-split Activation + Max-Weight Channel Scheduling:** We propose a solution that explicitly controls the time-scale separation between BS activation and rate allocation decisions. At BS switching instants (which occurs at a slow time-scale), the strategy uses a static-split rule (time-sharing) which is pre-computed using the explicit knowledge of the arrival and channel statistics for selecting the activation state. This activation algorithm is combined with a queue-length based Max-Weight algorithm for rate allocation (applied at the fast time-scale of channel variability). We show that the joint dynamics of these two algorithms leads to stability; further, the choice of parameters for the activation algorithm enables us to achieve an average network cost that can be made arbitrarily close to the optimal cost.
2) **Learning algorithm with provable guarantees:** In the setting where the arrival and channel statistics are not known, we propose an *explore-exploit* policy that estimates arrival and channel statistics in the explore phase, and uses the estimated statistics for activation decisions in the exploit phase (this phase includes BS switching at a slow time-scale). This is combined with a Max-Weight based rate allocation rule restricted to the activated BSs (at a fast time-scale). We prove that this joint learning-cum-scheduling algorithm can ensure queue stability while achieving close to optimal network cost.
3) **Convergence bounds for time-inhomogeneous Markov chains:** In the course of proving the theoretical guarantees for our algorithm, we derive useful technical results on convergence of time-inhomogeneous Markov chains. More specifically, we derive explicit convergence bounds for the marginal distribution of a finite-state time-inhomogeneous Markov chain whose transition probability matrices at each time-step are arbitrary (but small) perturbations of a given stochastic matrix. We believe that these bounds are useful not only in this specific problem, but are of independent interest.

To summarize then, our approach can be viewed as an algorithmically engineered separation of time-scales for only the activation set dynamics, while adapting to the channel variability for the queue dynamics. Such an engineering of time-scales leads to coupled fast-slow dynamics, the 'fast' due to opportunistic channel allocation and packet queue evolution with Max-Weight, and the 'slow' due to infrequent base-station switching using learned statistics. Through a novel Lyapunov technique for convergent time-inhomogeneous Markov chains, we show that we can achieve queue stability while operating at a near-optimal network cost.

Related Work

While mobile networks have been traditionally designed with the objective of optimizing spectral efficiency, design of energy efficient networks has been of recent interest. A survey of various techniques proposed to reduce operational costs and carbon footprint can be found in [3], [13], [2], [5]. The survey in [5] specially focuses on sleep mode techniques in BSs.

Various techniques have been proposed to exploit BS sleep mode to reduce energy consumption in different settings. Most of them aim to minimize energy consumption while guaranteeing minimum user QoS requirements. For example, [14], [6], [15] consider inelastic traffic and consider outage probability or blocking probability as metrics for measuring QoS. In [8], the problem is formulated as a utility optimization problem with the constraint that the minimum rate demand should be satisfied. But they do not explicitly evaluate the performance of their algorithm with respect to user QoS. The authors in [16], [17] model a single BS scenario with elastic traffic as an M/G/1 vacation queue and characterize the impact of sleeping on mean user delay and energy consumption. In [9], the authors consider the multi BS setting with Poisson arrivals and delay constraint at each BS.

Most papers that study BS switching use models that ignore switching cost. Nonetheless, a few papers acknowledge the importance of avoiding frequent switching. For example, Oh et al. [18] implement a hysteresis time for switching in their algorithm although they do not consider it in their theoretical analysis. Gou et al. [17] also study hysteresis sleeping schemes which enforce a minimum sleeping time. In [8] and [9], it is ensured that interval between switching times are large enough to avoid overhead due to transient network states. Finally Jie et al. [6] consider BS sleeping strategies which explicitly incorporate switching cost in the model (but they do not consider packet queue dynamics). They emphasize that frequent switching should be avoided considering its effect on signaling overhead, device lifetime and switching energy consumption, and also note that incorporating switching cost introduces time correlation in the system dynamics.

Notation: Important notation for the problem setting can be found in Table I. Boldface letters are used to denote vectors or matrices and the corresponding non-bold letters to denote their individual components. Also, the notation $\mathbb{1}\{\cdot\}$ is used to denote the indicator function. For any two vectors $\mathbf{v}_1$, $\mathbf{v}_2$ and scalar a, $\mathbf{v}_1 \cdot \mathbf{v}_2$ denotes the dot product between the two vectors and $\mathbf{v}_1 + a = \mathbf{v}_1 + a\mathbf{1}$.

TABLE I
GENERAL NOTATION

Symbol	Description
n	Number of users
M	Number of BSs
$A_{m,u}(t)$	Arrival for user u at BS m at time t
A	Maximum number of arrivals to any queue in a time-slot
$\boldsymbol{\lambda}$	Average arrival rate vector
$H(t)$	Channel state at time t
$\mathcal{H}$	Set of all possible channel states
$\boldsymbol{\mu}$	Probability mass function of channel state
R	Maximum service rate to any queue in a time-slot
$h\|_j$	Channel state h restricted to the activated BSs in j
$\mathcal{R}(j,h) \subseteq \mathbb{R}^{M\times n}$	Set of all possible rate vectors for activation vector j and channel state h
$\mathbf{J}(t) = (J_m(t))$	Activation vector at time t
$\mathcal{J}$	Set of all possible activation states
$\mathbf{S}(t) = (S_{m,u}(t))$	Rate allocation at time t
C_1	Cost of operating a BS in ON state
C_0	Cost of switching a BS from ON to OFF state
$C(t)$	Network cost at time t
$Q_{m,u}(t)$	Queue of user u at BS m at the beginning of time-slot t
$\mathcal{P}_l$	Set of all probability (row) vectors in $\mathbb{R}^l$
$\mathcal{P}_l^2$	Set of all stochastic matrices in $\mathbb{R}^{l\times l}$
$\mathcal{W}_l$	Set of all stochastic matrices in $\mathbb{R}^{l\times l}$ with a single ergodic class
$\mathbf{1}_l$	All 1's Column vector of size l
$\mathbf{I}_l$	Identity matrix of size l

II. System Model

We consider a time-slotted cellular network with n users and M base-stations (BS) indexed by $u = 1,\ldots,n$ and $m = 1,\ldots,M$ respectively. Users can possibly be connected to multiple BSs. It is assumed that the user-BS association does not vary with time.

A. Arrival and Channel Model

Data packets destined for a user u arrive at a connected BS m as a bounded (at most $\bar{\mathsf{A}}$ packets in any time-slot), i.i.d. process $\{A_{m,u}(t)\}_{t\geq 1}$ with rate $\mathbb{E}[A_{m,u}(t)] = \lambda_{m,u}$. Arrivals get queued if they are not immediately transmitted. Let $Q_{m,u}(t)$ represent the queue-length of user u at BS m at the beginning of time-slot t.

The channel between the BSs and their associated users is also time-varying and i.i.d across time (but can be correlated across links), which we represent by the network channel-state process $\{H(t)\}_{t>0}$. At any time t, $H(t)$ can take values from a finite set $\mathcal{H}$ with probability mass function given by $\boldsymbol{\mu}$. Let $\bar{\mathsf{R}}$ be the maximum number of packets that can be transmitted over any link in a single time-slot. We consider an abstract model for interference by defining the set $\mathcal{R}(\mathbf{1},h)$ as the set of all possible rate vectors when the channel state is h.

B. Resource Allocation

At any time-slot t, the scheduler has to make two types of allocation decisions:

BS Activation: Each BS can be scheduled to be in one of the two states, *ON* (*active* mode) and *OFF* (*sleep* mode). Packet transmissions can be scheduled only from BSs in the *ON* state. The cost of switching a BS from *ON* in the previous time-slot to *OFF* in the current time-slot is given by C_0 and the cost of maintaining a BS in the *ON* state in the current time-slot is given by C_1. The activation state at time t is denoted by $\mathbf{J}(t) = (J_m(t))_{m\in[M]}$, where $J_m(t) := \mathbb{1}\{\text{BS } m \text{ is } ON \text{ at time } t\}$. We also denote the set of all possible activation states by $\mathcal{J}$. The total cost of operation, which we refer to as the *network cost*, at time t is the sum of switching and activation cost and is given by

$$C(t) := \mathsf{C}_0\|(\mathbf{J}(t-1) - \mathbf{J}(t))^+\|_1 + \mathsf{C}_1\|\mathbf{J}(t)\|_1. \quad (1)$$

It is assumed that the current network channel-state $H(t)$ is unavailable to the scheduler at the time of making activation decisions.

Rate Allocation: The network channel-state is observed after the BSs are switched *ON* and before the packets are scheduled for transmission. Moreover, only the part of the channel state restricted to the activated BSs, which we denote by $H(t)|_{\mathbf{J}(t)}$, can be observed. For any $j \in \mathcal{J}, h \in \mathcal{H}$, let $\mathcal{R}(j,h) \subset \mathbb{R}^{M\times n}$ be the set of all possible service rate vectors that can be allocated when the activation set is j and the channel state is h. Given the channel observation $H(t)|_{\mathbf{J}(t)}$, the scheduler allocates a rate vector $\mathbf{S}(t) = (S_{m,u}(t))_{m\in[M],u\in[n]}$ from the set $\mathcal{R}(\mathbf{J}(t), H(t))$ for packet transmission. This allows for draining of $S_{m,u}(t)$ packets from user u's queue at BS m for all $u \in [n]$ and $m \in [M]$.

Thus the resource allocation decision in any time-slot t is given by the tuple $(\mathbf{J}(t), \mathbf{S}(t))$. The sequence of operations in any time-slot can, thus, be summarized as follows: (i) Arrivals, (ii) BS Activation-Deactivation, (iii) Channel Observation, (iv) Rate Allocation, and (v) Packet Transmissions.

C. Model Extensions

Some of the assumptions in the model above are made for ease of exposition and can be extended in the following ways without affecting the results in the paper:

(i) Network Cost: We assume that the cost of operating a BS in the *OFF* state (sleep mode) is zero. However, it is easy to include an additional parameter, say C_1', which denotes the cost of a BS in the *OFF* state. Similarly, for switching cost, although we consider only the cost of switching a BS from *ON* to *OFF* state, we can also include the cost of switching from *OFF* to *ON* state (say C_0'). The analysis in this paper can then be extended by defining the network cost as

$$C(t) = \mathsf{C}_0 \|(\mathbf{J}(t-1) - \mathbf{J}(t))^+\|_1 + \mathsf{C}_1 \|\mathbf{J}(t)\|_1 + \mathsf{C}_0' \|(\mathbf{J}(t) - \mathbf{J}(t-1))^+\|_1 + \mathsf{C}_1' \left(M - \|\mathbf{J}(t)\|_1\right)$$

instead of (1).

(ii) Switching Hysteresis Time: While our system allows switching decisions in every time-slot, we will see that the key to our approach is a slowing of activation set switching dynamics. Specifically, on average our algorithm switches activation states once every $1/\epsilon_s$ timeslots, where ϵ_s is a tunable parameter. Additionally, it is easy to incorporate "hard constraints" on the hysteresis time by restricting the frequency of switching decisions to, say once in every L time-slots (for some constant L). This avoids the problem of switching too frequently and gives a method to implement time-scale separation between the channel allocation decisions and BS activation decisions. While our current algorithm has inter-switching times i.i.d. geometric with mean $1/\epsilon_s$, it is easy to allow other distributions that have bounded means with some independence conditions (independent of each other and also the arrivals and the channel). We skip details in the proofs for notational clarity.

III. Optimization Framework

A policy is given by a (possibly random) sequence of resource allocation decisions $(\mathbf{J}(t), \mathbf{S}(t))_{t>0}$. Let $(\mathbf{J}(t-1), \mathbf{Q}(t))$ be the network state at time t.

Notation: We use $\mathbb{P}_\varphi[\cdot]$ and $\mathbb{E}_\varphi[\cdot]$ to denote probabilities and expectation under policy φ. We skip the subscript when the policy is clear from the context.

A. Stability, Network Cost, and the Optimization Problem

Definition 1 (Stability)**.** *A network is said to be* stable *under a policy φ if there exist constants $\bar{Q}$, $\rho > 0$ such that for any initial condition $(\mathbf{J}(0), \mathbf{Q}(1))$,*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{P}_\varphi \left[\sum_{m\in[m], u\in[n]} Q_{m,u}(t) \le \bar{Q} \,\middle|\, \mathbf{J}(0), \mathbf{Q}(1) \right] > \rho.$$

This definition is motivated by Foster's theorem: indeed, for an aperiodic and irreducible DTMC, Definition 1 implies positive recurrence. Consider the set of all ergodic Markov policies $\mathfrak{M}$, including those that know the arrival and channel statistics. A policy $\varphi \in \mathfrak{M}$ makes allocation decisions at time t based only on the current state $(\mathbf{J}(t-1), \mathbf{Q}(t))$ (and possibly the arrival and channel statistical parameters). We now define the support region of a policy and the capacity region.

Definition 2 (Support Region of a Policy φ)**.** *The support region $\Lambda^\varphi(\boldsymbol{\mu})$ of a policy φ is the set of all arrival rate vectors for which the network is stable under the policy φ.*

Definition 3 (Capacity Region)**.** *The capacity region $\Lambda(\boldsymbol{\mu})$ is the set of all arrival rate vectors for which the network is stable under some policy in $\mathfrak{M}$, i.e., $\Lambda(\boldsymbol{\mu}) := \bigcup_{\varphi\in\mathfrak{M}} \Lambda^\varphi(\boldsymbol{\mu})$.*

Definition 4 (Network Cost of a Policy φ)**.** *The network cost $C^\varphi(\boldsymbol{\mu}, \boldsymbol{\lambda})$ under a policy φ is the long term average network cost (BS switching and activation costs) per time-slot, i.e.,*

$$C^\varphi(\boldsymbol{\mu}, \boldsymbol{\lambda}) := \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_\varphi \left[C(t) \,\middle|\, \mathbf{J}(0), \mathbf{Q}(1) \right].$$

We formulate the resource allocation problem in a network cost minimization framework. Consider the problem of network cost minimization under Markov policies $\mathfrak{M}$ subject to stability. The optimal network cost is given by

$$C^{\mathfrak{M}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) := \inf_{\{\varphi\in\mathfrak{M}:\boldsymbol{\lambda}\in\Lambda^\varphi(\boldsymbol{\mu})\}} C^\varphi(\boldsymbol{\mu}, \boldsymbol{\lambda}). \quad (2)$$

B. Markov-Static-Split Rules

The capacity region $\Lambda(\boldsymbol{\mu})$ will naturally be characterized by only those Markov policies that maintain all the BSs active in all the time-slots, i.e., $\mathbf{J}(t) = \mathbf{1}\,\forall t$. In the traditional scheduling problem without BS switching, it is well-known that the capacity region can be characterized by the class of *static-split* policies [19] that allocate rates in a random i.i.d. fashion given the current channel state. An arrival rate vector $\boldsymbol{\lambda} \in \Lambda(\boldsymbol{\mu})$ iff there exists convex combinations $\left\{\boldsymbol{\alpha}(\mathbf{1}, h) \in \mathcal{P}_{|\mathcal{R}(\mathbf{1},h)|}\right\}_{h\in\mathcal{H}}$ such that

$$\boldsymbol{\lambda} < \sum_{h\in\mathcal{H}} \mu(h) \sum_{\mathbf{r}\in\mathcal{R}(\mathbf{1},h)} \alpha_{\mathbf{r}}(\mathbf{1}, h)\mathbf{r}.$$

But note that static-split rules in the above class, in which BSs are not switched *OFF*, do not optimize the network cost.

We now describe a class of activation policies called the *Markov-static-split + static-split* rules which are useful in handling the network cost. A policy is a Markov-static-split + static-split rule if it uses a time-homogeneous Markov rule for BS activation in every time-slot and an i.i.d. static-split rule for rate allocations. For any $l \in \mathbb{N}$, let $\mathcal{W}_l$ denote the set of all stochastic matrices of size l with a single ergodic class. A Markov-static-split + static-split rule is characterized by

1) a stochastic matrix $\mathbf{P} \in \mathcal{W}_{|\mathcal{J}|}$ with a single ergodic class,
2) convex combinations $\left\{\boldsymbol{\alpha}(j, h) \in \mathcal{P}_{|\mathcal{R}(j,h)|}\right\}_{j\in\mathcal{J}, h\in\mathcal{H}}$.

Here $\mathbf{P}$ represents the transition probability matrix that specifies the jump probabilities from one activation state to another in successive time-slots. $\{\boldsymbol{\alpha}(j, h)\}_{j\in\mathcal{J}, h\in\mathcal{H}}$ specify the static-split rate allocation policy given the activation state and the network channel-state.

Let $\mathfrak{MS}$ denote the class of all Markov-static-split + static-split rules. For a rule $\left(\mathbf{P}, \boldsymbol{\alpha} = \{\boldsymbol{\alpha}(j, h)\}_{j\in\mathcal{J}, h\in\mathcal{H}}\right) \in \mathfrak{MS}$, let $\boldsymbol{\sigma}$ denote the invariant probability distribution corresponding to the stochastic matrix $\mathbf{P}$. Then the expected switching and activation costs are given by $\mathsf{C}_0 \sum_{j',j\in\mathcal{J}} \sigma_{j'} P_{j',j} \|(j' - j)^+\|_1$

and $\mathsf{C}_1 \sum_{j\in\mathcal{J}} \sigma_j \|j\|_1$ respectively. We prove in the following theorem that the class $\mathfrak{MS}$ can achieve the same performance as $\mathfrak{M}$, the class of all ergodic Markov policies.

Theorem 1. *For any $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ and $\varphi \in \mathfrak{M}$ such that $\boldsymbol{\lambda} \in \Lambda^{\varphi}(\boldsymbol{\mu})$, there exists a $\varphi' \in \mathfrak{MS}$ such that $\boldsymbol{\lambda} \in \Lambda^{\varphi'}(\boldsymbol{\mu})$ and $C^{\varphi'}(\boldsymbol{\mu},\boldsymbol{\lambda}) = C^{\varphi}(\boldsymbol{\mu},\boldsymbol{\lambda})$. Therefore,*

$$C^{\mathfrak{M}}(\boldsymbol{\mu},\boldsymbol{\lambda}) = \inf_{\varphi'\in\mathfrak{MS},\lambda\in\Lambda^{\varphi'}(\boldsymbol{\mu})} C^{\varphi'}(\boldsymbol{\mu},\boldsymbol{\lambda}).$$

Proof Outline. The proof of this theorem is similar to the proof of characterization of the stability region using the class of static-split policies. It maps the time-averages of BS activation transitions and rate allocations of the policy $\varphi \in \mathfrak{M}$ to a Markov-static-split rule $\varphi' \in \mathfrak{MS}$ that mimics the same time-averages. The complete proof is available in [20]. □

From the characterization of the class $\mathfrak{MS}$, Theorem 1 shows that $C^{\mathfrak{M}}(\boldsymbol{\mu},\boldsymbol{\lambda})$ is equal to the optimal value, $V(\mu,\lambda)$, of the optimization problem (2), given by

$$\inf_{\mathbf{P},\boldsymbol{\alpha}} \mathsf{C}_0 \sum_{j',j\in\mathcal{J}} \sigma_{j'} P_{j',j} \|(j'-j)^+\|_1 + \mathsf{C}_1 \sum_{j\in\mathcal{J}} \sigma_j \|j\|_1$$

such that $\mathbf{P} \in \mathcal{W}_{|\mathcal{J}|}$ with unique invariant distribution $\boldsymbol{\sigma} \in \mathcal{P}_{|\mathcal{J}|}$, and $\boldsymbol{\alpha}(j,h) \in \mathcal{P}_{|\mathcal{R}(j,h)|}\, \forall j \in \mathcal{J}, h \in \mathcal{H}$ with

$$\boldsymbol{\lambda} < \sum_{j\in\mathcal{J}} \sigma_j \sum_{h\in\mathcal{H}} \mu(h) \sum_{\mathbf{r}\in\mathcal{R}(j,h)} \alpha_{\mathbf{r}}(j,h)\mathbf{r}. \tag{3}$$

C. A Modified Optimization Problem

Now, consider the linear program given by

$$\min_{\boldsymbol{\sigma},\boldsymbol{\beta}} \mathsf{C}_1 \sum_{j\in\mathcal{J}} \sigma_j \|j\|_1, \quad \text{such that}$$

$$\begin{aligned}
\boldsymbol{\sigma} &\in \mathcal{P}_{|\mathcal{J}|} \\
\beta_{j,h,\mathbf{r}} &\geq 0 \quad \forall \mathbf{r} \in \mathcal{R}(j,h), \forall j \in \mathcal{J}, h \in \mathcal{H}, \\
\sigma_j &= \sum_{\mathbf{r}\in\mathcal{R}(j,h)} \beta_{j,h,\mathbf{r}} \quad \forall j \in \mathcal{J}, h \in \mathcal{H}, \\
\boldsymbol{\lambda} &\leq \sum_{\substack{j\in\mathcal{J},h\in\mathcal{H},\\ \mathbf{r}\in\mathcal{R}(j,h)}} \beta_{j,h,\mathbf{r}} \mu(h)\mathbf{r}.
\end{aligned} \tag{4}$$

Let $d := |\mathcal{J}| + \sum_{j\in\mathcal{J},h\in\mathcal{H}} |\mathcal{R}(j,h)|$ be the number of variables in the above linear program. We denote by $L_{\mathbf{c}}(\boldsymbol{\mu},\boldsymbol{\lambda})$, a linear program with constraints as above and with $\mathbf{c} \in \mathbb{R}^d$ as the vector of weights in the objective function. Thus, the feasible set of the linear program $L_{\mathbf{c}}(\boldsymbol{\mu},\boldsymbol{\lambda})$ is specified by the parameters $\boldsymbol{\mu},\boldsymbol{\lambda}$ and the objective function is specified by the vector $\mathbf{c}$. Let $C^*_{\mathbf{c}}(\boldsymbol{\mu},\boldsymbol{\lambda})$ denote the optimal value of $L_{\mathbf{c}}(\boldsymbol{\mu},\boldsymbol{\lambda})$ and $\mathcal{O}^*_{\mathbf{c}}(\boldsymbol{\mu},\boldsymbol{\lambda})$ denote the optimal solution set. Also, let

$$\mathcal{S} := \{(\boldsymbol{\mu},\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda(\boldsymbol{\mu})\},$$

$$\mathcal{U}_{\mathbf{c}} := \{(\boldsymbol{\mu},\boldsymbol{\lambda}) \in \mathcal{S} : L_{\mathbf{c}}(\boldsymbol{\mu},\boldsymbol{\lambda}) \text{ has a unique solution}\}.$$

Note that $L_{\mathbf{c}^0}(\boldsymbol{\mu},\boldsymbol{\lambda})$, with

$$\mathbf{c}^0 := ((\mathsf{C}_1\|j\|_1)_{j\in\mathcal{J}}, \mathbf{0}) \tag{5}$$

is a relaxation of the original optimization problem $V(\boldsymbol{\mu},\boldsymbol{\lambda})$, and therefore

$$C^*_{\mathbf{c}}(\boldsymbol{\mu},\boldsymbol{\lambda}) \leq C^{\mathfrak{M}}(\boldsymbol{\mu},\boldsymbol{\lambda}). \tag{6}$$

We use results from [21], [22] to show (in the Lemma below) that the solution set and the optimal value of the linear program are continuous functions of the input parameters.

Lemma 1. *(I) As a function of the weight vector $\mathbf{c}$ and the parameters $\boldsymbol{\mu},\boldsymbol{\lambda}$, the optimal value $C^*_{(\cdot)}(\cdot)$ is continuous at any $(\mathbf{c},(\boldsymbol{\mu},\boldsymbol{\lambda})) \in \mathbb{R}^d \times \mathcal{S}$.*

(II) For any weight vector $\mathbf{c}$, the optimal solution set $\mathcal{O}^*_{\mathbf{c}}(\cdot)$, as a function of the parameters $(\boldsymbol{\mu},\boldsymbol{\lambda})$, is continuous at any $(\boldsymbol{\mu},\boldsymbol{\lambda}) \in \mathcal{U}_{\mathbf{c}}$.

Remark 1. *Since $\mathcal{O}^*_{\mathbf{c}}(\boldsymbol{\mu},\boldsymbol{\lambda})$ is a singleton if $(\boldsymbol{\mu},\boldsymbol{\lambda}) \in \mathcal{U}_{\mathbf{c}}$, the definition of continuity in this context is unambiguous.*

D. A Feasible Solution: Static-Split + Max-Weight

We now discuss how we can use the linear program L to obtain a feasible solution for the original optimization problem (2). We need to deal with two modified constraints:

(i) Single Ergodic Class – Spectral Gap: For any $\boldsymbol{\sigma} \in \mathcal{P}_{|\mathcal{J}|}$ and $\epsilon_s \in (0,1)$, the stochastic matrix

$$\mathbf{P}(\boldsymbol{\sigma},\epsilon_s) := \epsilon_s \mathbf{1}_{|\mathcal{J}|}\boldsymbol{\sigma} + (1-\epsilon_s)\mathbf{I}_{|\mathcal{J}|} \tag{7}$$

is aperiodic and has a single ergodic class given by $\{j : \sigma_j > 0\}$. Therefore, given any optimal solution $(\boldsymbol{\sigma},\boldsymbol{\beta})$ for the relaxed problem $L_{\mathbf{c}}(\boldsymbol{\mu},\boldsymbol{\lambda})$, we can construct a feasible solution $(\mathbf{P}(\boldsymbol{\sigma},\epsilon_s),\boldsymbol{\alpha})$ for the original optimization problem $V(\boldsymbol{\mu},\boldsymbol{\lambda})$ such that the network cost for this solution is at most $\epsilon_s M \mathsf{C}_0$ more than the optimal cost. Note that ϵ_s is the spectral gap of the matrix $\mathbf{P}(\boldsymbol{\sigma},\epsilon_s)$.

(ii) Stability – Capacity Gap: To ensure stability, it is necessary that the arrival rate is strictly less than the service rate (inequality (3)). It can be shown that an optimal solution to the linear program satisfies the constraint (4) with equality, and therefore cannot guarantee stability. An easy solution to this problem is to solve a modified linear program with a fixed small gap ϵ_g between the arrival rate and the offered service rate. We refer to the parameter ϵ_g as the *capacity gap*. Continuity of the optimal cost of the linear program L (from part (I) of Lemma 1) ensures that the optimal cost of the modified linear program is close to the optimal cost of the original optimization problem for sufficiently small ϵ_g.

To summarize, if the statistical parameters $\boldsymbol{\mu},\boldsymbol{\lambda}$ were known, one could adopt the following scheduling policy:

(a) BS activation: At every time-slot, with probability $1-\epsilon_s$, maintain the BSs in the same state as the previous time-slot, i.e., no switching. With probability ϵ_s, compute an optimal solution $(\boldsymbol{\sigma}^*,\boldsymbol{\beta}^*)$ for the linear program $L_{\mathbf{c}^0}(\boldsymbol{\mu},\boldsymbol{\lambda}+\epsilon_g)$ and schedule BS activation according to the static-split rule given by $\boldsymbol{\sigma}^*$. The network can be operated at a cost close to the optimal by choosing ϵ_s, ϵ_g sufficiently small.

(b) Rate allocation: To ensure stability, use a queue-based rule such as the Max-Weight rule to allocate rates given the observed channel state:

$$\mathbf{S}(t) = \underset{\mathbf{r}\in\mathcal{R}(\mathbf{J}(t),H(t))}{\arg\max}\ \mathbf{Q}(t)\cdot\mathbf{r}. \tag{8}$$

We denote the above static-split + Max-Weight rule with parameters ϵ_s, ϵ_g by $\varphi(\boldsymbol{\mu},\boldsymbol{\lambda}+\epsilon_g,\epsilon_s)$. The theorem below shows that the static-split + Max-Weight policy achieves close to optimal cost while ensuring queue stability.

Theorem 2. *For any $\boldsymbol{\mu},\boldsymbol{\lambda}$ such that $(\boldsymbol{\mu},\boldsymbol{\lambda}+2\epsilon_g)\in\mathcal{S}$, and for any $\epsilon_s\in(0,1)$, under the static-split + Max-Weight rule $\varphi(\boldsymbol{\mu},\boldsymbol{\lambda}+\epsilon_g,\epsilon_s)$,*

1) *the network cost satisfies*

$$C^{\varphi(\boldsymbol{\mu},\boldsymbol{\lambda}+\epsilon_g,\epsilon_s)}(\boldsymbol{\mu},\boldsymbol{\lambda})\le C^{\mathfrak{M}}(\boldsymbol{\mu},\boldsymbol{\lambda})+\kappa\epsilon_s+\gamma(\epsilon_g),$$

for some constant κ that depends on the network size and C_0, C_1, and for some increasing function $\gamma(\cdot)$ such that $\lim_{\epsilon_g\to 0}\gamma(\epsilon_g)=0$, and

2) *the network is stable, i.e.,*

$$\boldsymbol{\lambda}\in\Lambda^{\varphi(\boldsymbol{\mu},\boldsymbol{\lambda}+\epsilon_g,\epsilon_s)}(\boldsymbol{\mu}).$$

Proof Outline. Since $\mathbf{P}(\boldsymbol{\sigma}^*,\epsilon_s)$ has a single ergodic class, the marginal distribution of the activation state $(\mathbf{J}(t))_{t>0}$ converges to $\boldsymbol{\sigma}^*$. Part 1 of the theorem then follows from (6) and the continuity of the optimal value of L (Lemma 1(I)). Part 2 relies on the strict inequality gap enforced by ϵ_g in (3). Therefore, it is possible to serve all the arrivals in the long-term. We use a standard Lyapunov argument which shows that the T-step quadratic Lyapunov drift for the queues is strictly negative outside a finite set for some $T>0$. The complete proof is available in [20]. □

One can also achieve the above guarantees with a static-split + static-split rule which has BS activations as above but channel allocation through a static-split rule with convex combinations given by $\boldsymbol{\alpha}^*$ such that

$$\alpha^*_{\mathbf{r}}(j,h)=\frac{\beta^*_{j,h,\mathbf{r}}}{\sigma^*_j}\quad\forall\mathbf{r}\in\mathcal{R}(j,h),\forall j\in\mathcal{J},h\in\mathcal{H}. \tag{9}$$

E. Effect of Parameter Choice on Performance

ϵ_s and ϵ_g can be used as a control parameters to trade-off between two desirable but conflicting features – small queue lengths and low network cost.

(i) Spectral gap, ϵ_s: ϵ_s is the spectral gap of the transition probability matrix $\mathbf{P}(\boldsymbol{\sigma}^*,\epsilon_s)$ and, therefore, impacts the mixing time of the activation state $(\mathbf{J}(t))_{t>0}$. Since the average available service rate is dependent on the distribution of the activation state, the time taken for the queues to stabilize depends on the mixing time, and consequently, on the choice of ϵ_s. With $\epsilon_s=1$, we are effectively ignoring switching costs as this corresponds to a rule that chooses the activation sets in an i.i.d. manner according to the distribution $\boldsymbol{\sigma}^*$. Thus, stability is ensured but at a penalty of larger average costs. At the other extreme, when $\epsilon_s=0$, the transition probability matrix $\mathbf{I}_{|\mathcal{J}|}$ corresponds to an activation rule that never switches the BSs from their initial activation state. This extreme naturally achieves zero switching cost, but at the cost of queue stability as the activation set is frozen for all times.

(ii) Capacity gap, ϵ_g: Recall that ϵ_g is the gap enforced between the arrival rate and the allocated service rate in the linear program $L_{\mathbf{c}^0}(\boldsymbol{\mu},\boldsymbol{\lambda}+\epsilon_g)$. Since the mean queue-length is known to vary inversely as the capacity gap, the parameter ϵ_g can be used to control queue-lengths. A small ϵ_g results in low network cost and large mean queue-lengths.

IV. Policy with Unknown Statistics

In the setting where arrival and channel statistics are unknown, our interest is in designing policies that learn the arrival and channel statistics to make rate allocation and BS activation decisions. As described in Section II-B, channel rates are observed in every time-slot after activation of the BSs. Since only channel rates of activated BSs can be obtained in any time-slot, the problem naturally involves a trade-off between activating more BSs to get better channel estimates versus maintaining low network cost. Our objective is to design policies that achieve network cost close to $C^{\mathfrak{M}}$ while learning the statistics well enough to stabilize the queues.

A. An Explore-Exploit Policy

Algorithm 1 gives a policy $\phi(\epsilon_p,\epsilon_s,\epsilon_g)$, which is an explore-exploit strategy similar to the ϵ-greedy policy in the multi-armed bandit problem. Here, $\epsilon_p,\epsilon_s,\epsilon_g$ are fixed parameters of the policy.

1) Initial Perturbation of the Cost Vector: Given the original cost vector $\mathbf{c}^0$ (given by (5)), the policy first generates a slightly perturbed cost vector $\mathbf{c}^{\epsilon_p}$ by adding to $\mathbf{c}^0$, a random perturbation uniformly distributed on the ϵ_p-ball. It is easily verified that, for any $(\boldsymbol{\mu},\boldsymbol{\lambda})\in\mathcal{S}$,

$$|C^*_{\mathbf{c}^{\epsilon_p}}(\boldsymbol{\mu},\boldsymbol{\lambda})-C^*_{\mathbf{c}^0}(\boldsymbol{\mu},\boldsymbol{\lambda})|\le\sqrt{|\mathcal{H}|+1}\mathsf{C}_1\epsilon_p.$$

In addition, the following lemma shows that the perturbed linear program has a unique solution with probability 1.

Lemma 2. *For any $(\boldsymbol{\mu},\boldsymbol{\lambda})\in\mathcal{S}$,*

$$\mathbb{P}\left[(\boldsymbol{\mu},\boldsymbol{\lambda})\in\mathcal{U}_{\mathbf{c}^{\epsilon_p}}\mid\mathbf{J}(0),\mathbf{Q}(1)\right]=1.$$

2) BS Activation: At any time t the policy randomly chooses to *explore* or *exploit*. The probability that it explores, $\epsilon_l(t)=\frac{2\log t}{t}$, decreases with time.

Exploration. In the *explore* phase, the policy activates all the BSs and observes the channel. It maintains $\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\lambda}}$, the empirical distribution of the channel and the empirical mean of the arrival vector respectively, obtained from samples in the explore phase.

Exploitation. In the *exploit* phase, with probability $1-\epsilon_s$, the policy chooses to keep the same activation set as the previous time-slot (i.e., no switching). With probability ϵ_s, it solves the linear program $L_{\mathbf{c}^{\epsilon_p}}\left(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\lambda}}+\epsilon_g\right)$ with the perturbed cost vector $\mathbf{c}^{\epsilon_p}$ and parameters $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\lambda}}+\epsilon_g$ given by the empirical distribution. From an optimal solution $\left(\hat{\boldsymbol{\sigma}}(t),\hat{\boldsymbol{\beta}}(t)\right)$ of the linear program, it chooses the BS activation vector $\mathbf{J}(t)$ according to the distribution $\hat{\boldsymbol{\sigma}}(t)$.

Algorithm 1 Policy $\phi(\epsilon_p, \epsilon_s, \epsilon_g)$ with parameters ϵ_p, ϵ_s, ϵ_g

Generate a uniformly distributed random direction $\boldsymbol{\upsilon} \in \mathbb{R}^d$.
Construct a perturbed weight vector

$$\mathbf{c}^{\epsilon_p} \leftarrow \mathbf{c}^0 + \epsilon_p \boldsymbol{\upsilon}.$$

Initialize $\hat{\boldsymbol{\mu}} \leftarrow \mathbf{0}$, $\hat{\boldsymbol{\lambda}} \leftarrow \mathbf{0}$.
for all $t > 0$ **do**
  Generate $E_l(t)$, an independent Bernoulli sample of mean $\epsilon_l(t) = \frac{2 \log t}{t}$.
  if $E_l(t) = 1$ **then** ▷ *Explore*
    $\mathbf{J}(t) \leftarrow \mathbf{1}$ (Activate all the BSs).
    Observe the channel state $H(t)$.
    Update empirical distributions $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\lambda}}$.
  else ▷ *Exploit*
    Generate $E_s(t)$, an independent Bernoulli sample of mean ϵ_s.
    if $E_s(t) = 0$ **then** ▷ *No Switching*
      $\mathbf{J}(t) \leftarrow \mathbf{J}(t-1)$.
    else
      Solve $L_{\mathbf{c}^{\epsilon_p}}\left(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\lambda}} + \epsilon_g\right)$ and select an optimal solution $\left(\hat{\boldsymbol{\sigma}}(t), \hat{\boldsymbol{\beta}}(t)\right)$.
      Select $\mathbf{J}(t)$ according to the distribution $\hat{\boldsymbol{\sigma}}(t)$.
    end if
    Observe the channel state $H(t)|_{\mathbf{J}(t)}$.
  end if
  Allocate channels according to the Max-Weight Rule,

$$\mathbf{S}(t) \leftarrow \underset{\mathbf{r} \in \mathcal{R}(\mathbf{J}(t), H(t))}{\arg\max} \mathbf{Q}(t) \cdot \mathbf{r}.$$

end for

3) Rate Allocation: The policy uses the Max-Weight Rule given by (8) for channel allocation.

B. Performance Guarantees

In Theorem 3, we give stability and network cost guarantees for the proposed learning-cum-scheduling rule $\phi(\epsilon_p, \epsilon_s, \epsilon_g)$.

Theorem 3. *For any $\boldsymbol{\mu}, \boldsymbol{\lambda}$ such that $(\boldsymbol{\mu}, \boldsymbol{\lambda} + 2\epsilon_g) \in \mathcal{S}$, and for any $\epsilon_p, \epsilon_s \in (0,1)$, under the policy $\phi(\epsilon_p, \epsilon_s, \epsilon_g)$,*

1) *the network cost satisfies*

$$C^{\phi(\epsilon_p, \epsilon_s, \epsilon_g)}(\boldsymbol{\mu}, \boldsymbol{\lambda}) \leq C^{\mathfrak{M}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) + \kappa(\epsilon_p + \epsilon_s) + \gamma(\epsilon_g),$$

for some constant κ that depends on the network size and C_0, C_1, and for some increasing function $\gamma(\cdot)$ such that $\lim_{\epsilon_g \to 0} \gamma(\epsilon_g) = 0$, and

2) *the network is stable, i.e.,*

$$\boldsymbol{\lambda} \in \Lambda^{\phi(\epsilon_p, \epsilon_s, \epsilon_g)}(\boldsymbol{\mu}).$$

Proof Outline. As opposed to known statistical parameters for the arrivals and the channel in the Markov-static-split rule, the policy uses empirical statistics that change dynamically with time. Thus, the activation state process $(\mathbf{J}(t))_{t>0}$, in this case, is not a time-homogeneous Markov chain. However, we note that $\mathbf{J}(t)$ along with the empirical statistics forms a time-inhomogeneous Markov chain with the empirical statistics converging to the true statistics almost surely. Specifically, we show that the time taken by the algorithm to learn the parameters within a small error has a finite second moment.

We then use convergence results for time-inhomogeneous Markov chains (derived in Lemma 3 in Section V) to show convergence of the marginal distribution of the activation state $(\mathbf{J}(t))_{t>0}$. As in Theorem 2, Part 1 then follows from (6) and the continuity of the optimal value of L (Lemma 1(I)).

Part 2 requires further arguments. The queues have a negative Lyapunov drift only after the empirical estimates have converged to the true parameters within a small error. To bound the Lyapunov drift before this time, we use boundedness of the arrivals along with the existence of second moment for the convergence time of the estimated parameters. By using a telescoping argument as in Foster's theorem, we show that this implies stability as per Definition 1. The complete proof is available in [20]. □

C. Discussion: Other Potential Approaches

Recall that our system consists of two distinct time-scales: (a) exogenous fast dynamics due the channel variability, that occurs on a per-time-slot basis, and (b) endogenous slow dynamics of learning and activation due to base-station active-sleep state dynamics. By 'exogenous', we mean that the time-scale is controlled by nature (channel process), and by 'endogenous', we mean that the time-scale is controlled by the learning-cum-activation algorithm (slowed dynamics where activation states change only infrequently). To place this in perspective, consider the following alternate approaches, each of which has defects.

1. Virtual queues + MaxWeight: As is now standard [10], [12], suppose that we encode the various costs through virtual queues (or variants there-of), and apply a MaxWeight algorithm to this collection of queues. Due to the switching cost, effective channel – the vector of channel rates on the active collection of base-stations – has dependence across time (coupled dynamics of channel and queues) through the activation set scheduling, and voids the standard Lyapunov proof approach for showing stability. Specifically, we cannot guarantee that the time average of various activation sets chosen by this (virtual + actual queue) MaxWeight algorithm equals the corresponding optimal fractions computed using a linear program with known channel and arrival parameters.

2. Ignoring Switching Costs with Fast Dynamics: Suppose we use virtual queues to capture only the activation costs. In this case, a MaxWeight approach (selecting a new activation set and channel allocation in each time-slot) will ensure stability, but will not provide any guarantees on cost optimality as there will be frequent switching of the activation set.

3. Ignoring Switching Costs with Slowed Dynamics: Again, we use virtual queues for encoding only activation costs, and use block scheduling. In other words, re-compute an activation + channel schedule once every R time-slots, and use this fixed schedule for this block of time (pick-and-compare, periodic, frame-based algorithms [23], [24], [25], [26]). While this approach minimizes switching costs (as activation changes occur infrequently), stability properties are lost as we are not making use of opportunism arising from the wireless channel variability (the schedule is fixed for a block of time and does not adapt to instantaneous channel variations).

Our approach avoids the difficulties in each of these approaches by explicitly slowing down the time-scale of the activation set dynamics (an engineered slow time-scale), thus minimizing switching costs. However, it allows channels to be opportunistically re-allocated in each time-slot based on the instantaneous channel state (the fast time-scale of nature). This fast-slow co-evolution of learning, activation sets and queue lengths requires a new proof approach. We combine new results (see Section V) on convergence of inhomogeneous Markov chains with Lyapunov analysis to show both stability and cost (near) optimality.

V. Convergence of a Time-Inhomogeneous Markov Process

We now derive some convergence bounds for perturbed time-inhomogeneous Markov chains which are useful in proving stability and cost optimality. Let $\mathcal{P} := \{\mathbf{P}_\delta, \delta \in \Delta\}$ be a collection of stochastic matrices in $\mathbb{R}^{N\times N}$, with $\{\boldsymbol{\sigma}_\delta, \delta \in \Delta\}$ denoting the corresponding invariant probability distributions. Also, let $\mathbf{P}_*$ be an $N \times N$ aperiodic stochastic matrix with a single ergodic class and invariant probability distribution $\boldsymbol{\sigma}_*$. Recall that for a stochastic matrix $\mathbf{P}$ the coefficient of ergodicity [27] $\tau_1(\mathbf{P})$ is defined by

$$\tau_1(\mathbf{P}) \;:=\; \max_{\mathbf{z}^\mathsf{T}\mathbf{1}_N=0\,,\,\|\mathbf{z}\|_1=1} \|\mathbf{P}^\mathsf{T}\mathbf{z}\|_1. \tag{10}$$

It has the following basic properties [27]:

1) $\tau_1(\mathbf{P}_1\mathbf{P}_2) \;\le\; \tau_1(\mathbf{P}_1)\tau_1(\mathbf{P}_2)$,
2) $|\tau_1(\mathbf{P}_1) - \tau_1(\mathbf{P}_2)| \;\le\; \|\mathbf{P}_1 - \mathbf{P}_2\|_\infty$,
3) $\|\mathbf{x}\mathbf{P} - \mathbf{y}\mathbf{P}\|_1 \;\le\; \tau_1(\mathbf{P})\,\|\mathbf{x}-\mathbf{y}\|_1 \quad \forall \mathbf{x},\mathbf{y} \in \mathcal{P}_N$, and
4) $\tau_1(\mathbf{P}) < 1$ if and only if $\mathbf{P}$ has no pair of orthogonal rows (i.e., if it is a scrambling matrix).

From [28], if $\mathbf{P}_*$ is aperiodic and has a single ergodic class then there exists an integer $\hat{m}$ such that $\mathbf{P}_*^k$ is scrambling for all $k \ge \hat{m}$. Therefore, $\tau_1(\mathbf{P}_*^k) < 1 \; \forall k \ge \hat{m}$. Define

$$\epsilon := \sup_{\delta\in\Delta} \|\mathbf{P}_\delta - \mathbf{P}_*\|_1. \tag{11}$$

Now, consider a time-inhomogeneous Markov chain $(X(t))_{t\ge 0}$ with initial distribution $\mathbf{y}(0)$, and transition probability matrix at time t given by $\mathbf{P}_{\delta_t} \in \mathcal{P} \; \forall t > 0$. Let $\{\mathbf{y}(t)\}_{t\ge0}$ be the resulting sequence of marginal distributions. The following lemma gives a bound on the convergence of the limiting distribution of such a time-inhomogeneous DTMC to $\boldsymbol{\sigma}_*$. Additional results are available in the technical report [20].

Lemma 3. *For any* $\mathbf{y}(0)$*,*

(a) the marginal distribution satisfies

$$\|\mathbf{y}(n) - \boldsymbol{\sigma}_*\|_1 \;\le\; \tau_1(\mathbf{P}_*^n)\|\mathbf{y}(0) - \boldsymbol{\sigma}_*\|_1 + \epsilon \sum_{\ell=0}^{n-1} \tau_1(\mathbf{P}_*^\ell)\,, \tag{12}$$

(b) and the limiting distribution satisfies

$$\limsup_{n\to\infty} \|\mathbf{y}(n) - \boldsymbol{\sigma}_*\|_1 \;\le\; \epsilon\,\Upsilon(\mathbf{P}_*)$$

where $\Upsilon(\mathbf{P}_*) := \sum_{\ell=0}^\infty \tau_1(\mathbf{P}_*^\ell) \le \frac{\hat{m}}{1-\tau_1(\mathbf{P}_*^{\hat{m}})}$.

Proof. The trajectory $(\mathbf{y}(n))_{n>0}$ satisfies $\forall n \ge 1$,

$$\mathbf{y}(n) \;=\; \mathbf{y}(n-1)\mathbf{P}_* + \mathbf{y}(n-1)(\mathbf{P}_{\delta_{n-1}} - \mathbf{P}_*). \tag{13}$$

Using (13) recursively, we have

$$\mathbf{y}(n) \;=\; \mathbf{y}(0)\mathbf{P}_*^n + \sum_{k=1}^{n} \mathbf{y}(n-k)(\mathbf{P}_{\delta_{n-k}} - \mathbf{P}_*)\mathbf{P}_*^{k-1},$$

which gives us

$$\mathbf{y}(n) - \boldsymbol{\sigma}_* = (\mathbf{y}(0) - \boldsymbol{\sigma}_*)\mathbf{P}_*^n + \sum_{k=1}^{n} \mathbf{y}(n-k)(\mathbf{P}_{\delta_{n-k}} - \mathbf{P}_*)\mathbf{P}_*^{k-1}. \tag{14}$$

Now, taking norms and using the definitions in (10) and (11), we obtain

$$\|\mathbf{y}(n) - \boldsymbol{\sigma}_*\|_1 \;\le\; \tau_1(\mathbf{P}_*^n)\|\mathbf{y}(0) - \boldsymbol{\sigma}_*\|_1 + \epsilon \sum_{\ell=0}^{n-1} \tau_1(\mathbf{P}_*^\ell)\,.$$

This proves part (a) of the lemma. Now, note that

$$\tau_1(\mathbf{P}_*^k) \;\le\; \left(\tau_1(\mathbf{P}_*^m)\right)^{\lfloor k/m \rfloor} \tag{15}$$

for any positive integers k, m. Since $\tau_1(\mathbf{P}_*^{\hat{m}}) < 1$, it follows that $\lim_{n\to\infty} \tau_1(\mathbf{P}_*^n) = 0$, and

$$\Upsilon(\mathbf{P}_*) \;=\; \sum_{\ell=0}^{\infty} \tau_1(\mathbf{P}_*^\ell) \;\le\; \frac{\hat{m}}{1 - \tau_1(\mathbf{P}_*^{\hat{m}})}.$$

Using this in (12), we have part (b). □

VI. Simulation Results

We present simulations that corroborate the results of this paper. The setting is as follows. There are five users and three BSs in the system. BS 1 can service users 1, 2, and 5. BS 2 can service users 1, 2, 3, and 4. BS 3 can service users 3, 4, and 5. The Bernoulli arrival rates on each queue (which have to be learned by the algorithm) is 0.1 packets/slot on each mobile-BS service connection. The total arrival rate to the system is thus 0.1 packet/slot × 10 connections, or 1 packet/slot. A good channel yields a service of 2 packets/slot while a bad channel yields 1 packet/slot. In our correlated fading model, either all channels are bad, or all connections to exactly one BS are good while the others bad. This yields four correlated channel states and all four are equiprobable (the probabilities being unknown to the algorithm). The fading process is independent and identically distributed over time. The activation constraint is that each BS can service at most one mobile per slot. The per BS switching cost C_0 and activation cost C_1 are both taken to be 1. Figure 1 provides the average queue sizes (first two plots) and average costs (third plot) for two values of ϵ_s, namely, 0.2 (first plot) and 0.05 (second plot). The plots show that a smaller ϵ_s yields a lower average cost and stabilizes the queue, but has higher average queue size.

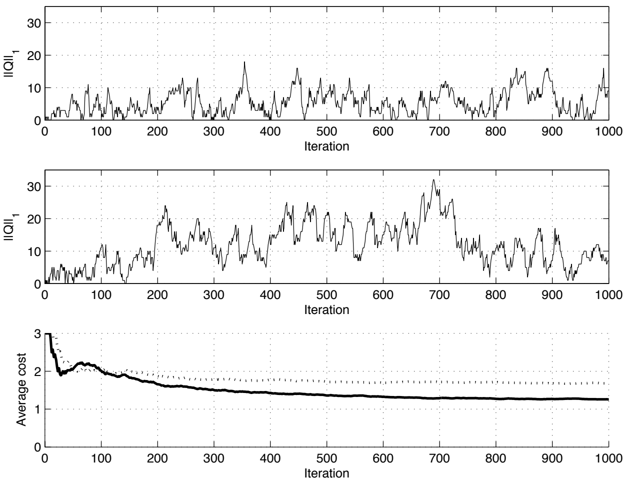

Fig. 1. The top two plots show the total queue size as a function of time when $\epsilon_s = 0.2$ and $\epsilon_s = 0.05$, respectively. The bottom plot shows the corresponding average costs (with the solid curve for $\epsilon_s = 0.05$). A smaller ϵ_s yields a lower average cost but has higher average queue size.

VII. Conclusion

We study the problem of jointly activating base-stations along with channel allocation, with the objective of minimizing energy costs (activation + switching) subject to packet queue stability. Our approach is based on timescale decomposition, consisting of fast-slow co-evolution of user queues (fast) and base-station activation sets (slow). We develop a learning-cum-scheduling algorithm that can achieve an average cost that is arbitrarily close to optimal, and simultaneously stabilize the user queues.

Acknowledgements

This work was partially supported by NSF grants CNS-1017549, CNS-1161868 and CNS-1343383, Army Research Office W911NF-17-1-0019, the US DoT supported D-STOP Tier 1 University Transportation Center, and the Robert Bosch Centre for Cyber Physical Systems.

References

[1] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, and A. Damnjanovic, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 82–89, 2014.

[2] G. Wu, C. Yang, S. Li, and G. Y. Li, "Recent advances in energy-efficient networks and their application in 5g systems," *IEEE Wireless Communications*, vol. 22, no. 2, pp. 145–151, 2015.

[3] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 56–61, 2011.

[4] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy savings in cellular access networks," in *2009 IEEE International Conference on Communications Workshops*. IEEE, 2009, pp. 1–5.

[5] J. Wu, Y. Zhang, M. Zukerman, and E. K.-N. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 803–826, 2015.

[6] G. Jie, Z. Sheng, and N. Zhisheng, "A dynamic programming approach for base station sleeping in cellular networks," *IEICE transactions on communications*, vol. 95, no. 2, pp. 551–562, 2012.

[7] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks," in *2010 Future Network & Mobile Summit*. IEEE, 2010.

[8] A. Abbasi and M. Ghaderi, "Distributed base station activation for energy-efficient operation of cellular networks," in *Proceedings of the 16th ACM international conference on Modeling, analysis & simulation of wireless and mobile systems*. ACM, 2013, pp. 427–436.

[9] J. Zheng, Y. Cai, X. Chen, R. Li, and H. Zhang, "Optimal base station sleeping in green cellular networks: A distributed cooperative framework based on game theory," *IEEE Transactions on Wireless Communications*, vol. 14, no. 8, pp. 4391–4406, 2015.

[10] L. Georgiadis, M. J. Neely, and L. Tassiulas, *Resource Allocation and Cross-Layer Control in Wireless Networks*. NOW Publishers, Foundations and Trends in Networking, 2006.

[11] X. Lin, N. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE Journal on Selected Areas in Comm.*, 2006.

[12] R. Srikant and L. Ying, *Communication Networks – An Optimization, Control, and Stochastic Networks Perspective*. Cambridge University Press, 2014.

[13] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Communications surveys & tutorials*, vol. 13, no. 4, pp. 524–540, 2011.

[14] F. Han, Z. Safar, W. S. Lin, Y. Chen, and K. R. Liu, "Energy-efficient cellular network operation via base station cooperation," in *2012 IEEE International Conference on Communications (ICC)*. IEEE, 2012, pp. 4374–4378.

[15] J. Gong, J. S. Thompson, S. Zhou, and Z. Niu, "Base station sleeping and resource allocation in renewable energy powered cellular networks," *IEEE Trans. on Communications*, vol. 62, no. 11, pp. 3801–3813, 2014.

[16] I. Kamitsos, L. Andrew, H. Kim, and M. Chiang, "Optimal sleep patterns for serving delay-tolerant jobs," in *Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking*. ACM, 2010, pp. 31–40.

[17] X. Guo, Z. Niu, S. Zhou, and P. Kumar, "Delay-constrained energy-optimal base station sleeping control," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1073–1085, 2016.

[18] E. Oh, K. Son, and B. Krishnamachari, "Dynamic base station switching-on/off strategies for green cellular networks," *IEEE transactions on wireless communications*, vol. 12, no. 5, pp. 2126–2136, 2013.

[19] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "CDMA data QoS scheduling on the forward link with variable channel conditions," *Bell Labs Tech. Memo*, April 2000.

[20] S. Krishnasamy, P. T. Akhil, A. Arapostathis, S. Shakkottai, and R. Sundaresan, "Augmenting max-weight with explicit learning for wireless scheduling with switching costs," tech. Report, UT Austin, January 2017.

[21] R. J. B. Wets, "On the continuity of the value of a linear program and of related polyhedral-valued multifunctions," *Mathematical programming study*, no. 24, pp. 14–29, 1985.

[22] M. Davidson, "Stability of the extreme point set of a polyhedron," *Journal of optimization theory and applications*, vol. 90, no. 2, pp. 357–380, 1996.

[23] L. Tassiulas, "Linear complexity algorithms for maximum throughput in radio networks and input queued switches," in *IEEE Infocom*, 1998.

[24] M. J. Neely, E. Modiano, and C. E. Rohrs, "Tradeoffs in delay guarantees and computation complexity for n n packet switches," in *Proceedings of CISS*, 2002.

[25] P. Chaporkar and S. Sarkar, "Stable scheduling policies for maximizing throughput in generalized constrained queueing," in *IEEE Infocom*, 2006.

[26] Y. Yi, A. Proutiere, and M. Chiang, "Complexity in wireless scheduling: Impact and tradeoffs," in *Proc. of the 9th ACM International Symp. on Mobile Ad Hoc Networking and Computing (MobiHoc)*, 2008.

[27] E. Seneta, *Non-negative matrices and Markov chains*. Springer Science & Business Media, 2006.

[28] J. M. Anthonisse and H. Tijms, "Exponential convergence of products of stochastic matrices," *Journal of Mathematical Analysis and Applications*, vol. 59, no. 2, pp. 360–364, 1977.