Self-Organizing Relays: Dimensioning, Self-Optimization and Learning

Richard Combes*, Zwi Altman* and Eitan Altman†
*Orange Labs
38/40 rue du Général Leclerc,92794 Issy-les-Moulineaux
Email:{richard.combes,zwi.altman}@orange.com
†INRIA Sophia Antipolis
06902 Sophia Antipolis, France
Email:Eitan.Altman@sophia.inria.fr

Abstract

Relay stations are an important component of heterogeneous networks introduced in the LTE-Advanced technology as a means to provide very high capacity and QoS all over the cell area. This paper develops a self-organizing network (SON) feature to optimally allocate resources between backhaul and station to mobile links. Static and dynamic resource sharing mechanisms are investigated. For stationary ergodic traffic we provide a queuing model to calculate the optimal resource sharing strategy and the maximal capacity of the network analytically. When traffic is not stationary, we propose a load balancing algorithm to adapt both the resource sharing and the zones covered by the relays based on measurements. Convergence to an optimal configuration is proven using stochastic approximation techniques. Self-optimizing dynamic resource allocation is tackled using a Markov Decision Process model. Stability in the infinite buffer case and blocking rate and file transfer time in the finite buffer case are considered. For a scalable solution with a large number of relays, a well-chosen parameterized family of policies is considered, to be used as expert knowledge. Finally, a model-free approach is shown in which the network can derive the optimal parameterized policy, and the convergence to a local optimum is proven. ([1],[2])

Index Terms

Relay, Queuing Theory, Stochastic Approximation, Reinforcement Learning, Stability, OFDMA, Load Balancing, Self configuration, Self Optimization

I. Introduction

Self-organizing networks (SON) mechanisms have been introduced in the Long Term Evolution (LTE) standard in order to empower the network by embedding autonomic mechanisms, namely self-configuration, self-optimization and self-healing ([1], [2]). These mechanisms aim at simplifying the network management, at reducing its cost of operation and at increasing its performance.

Dynamic self-optimization targets on-line network implementation of SON mechanisms with short time resolution (e.g. seconds to minutes) for adapting the network to new operation conditions such as traffic variations. The requirements for SON solutions to be adopted in radio access networks are the classical goodness criteria in optimization and control: existence of optimal solutions, convergence to an optimal solution, speed of convergence, monotonic improvement of the goodness of the solution, stability and robustness to noise. Previous works on on-line network optimization include the popular utility-based approach used in [3], [4] and [5]. Reinforcement learning has been investigated for example in [6].

LTE-Advanced introduces the concept of Heterogeneous Network (HetNet) as a means to increase network capacity. HetNets comprise low power nodes deployed in high traffic areas to increase capacity, namely picocells, femtocells and Relay Stations (RSs). Autonomous resource management in HetNets is among the important and challenging research avenues in SON for next generation radio access networks, encompassing load balancing,

[1] Manuscript received February 20, 2012; revised June 08, 2012; approved by IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT associate editor Athanasios Vasilakos.

[2] This work has been partially carried out in the framework of the FP7 UniverSelf project under EC Grant agreement 257513

Inter-Cell Interference Coordination (ICIC), mobility management, and other self-optimizing resource allocation mechanisms.

This paper focuses on self-optimizing RSs. RSs are linked to the macrocell by a wireless link which replaces the wired backhaul. We will use the term "station" to refer to a Base Station (BS) or a RS indifferently. Radio resources have to be shared between the BS to RSs links and the stations to users' links. The resource allocation which maximizes the system capacity depends on system parameters such as traffic and RSs placement. Both static and dynamic mechanisms are investigated in this work.

We first derive the static resource allocation which maximizes the system capacity. We then show a dynamic resource allocation as an optimal control problem. We give a systematic method for the controller design, in three steps:

1) The problem is modelled as a Markov Decision Process (MDP), and the optimal controller is found. This optimal controller is to be used as expert knowledge during the next phase.
2) Based on the previous controller and a queuing theory result, we introduce a set of parameterized policies (the expert knowledge). A method to find the optimal parameterized controller is derived and its performance is compared with the optimal controller.
3) Finally, we show a model-free (reinforcement learning) approach to derive the optimal parameterized policy by observation and interaction with the network. We use the policy-gradient method featured in [7], [8], [9].

The contributions of the present paper are:

1) A queuing analysis to derive the optimal static resource allocation in closed form, and the impact of the major system parameters such as RS placement, number of deployed RSs and RS size on the system performance.
2) A self-organizing algorithm to adapt the network to traffic variations automatically. Both the zones covered by RSs and the resources allocated to the backhaul are adapted simultaneously and convergence to an optimal configuration is proven using stochastic approximation.
3) A systematic step-by-step framework for controller design, with rigorous proofs of convergence and optimality of the methods used.
4) A model-free approach with monotonic improvement of the solution during the learning phase. This is fundamental for on-line implementation in an operational network.

The paper is organized as follows: Section II states the system model and the optimal static resource allocation strategy is derived in closed form based on a queuing analysis. The impact of RS placement, number of deployed RS and RS size is investigated. In section III, we show that the network can adapt itself to traffic variations based on traffic measurements, allowing automatic traffic balancing. Section IV models the problem as a MDP, and a parameterized set of policies is derived based on the optimal policy. Section V presents a model-free approach to derive the optimal parameterized policy by interaction with the network, without degradation during the learning phase. Section VI concludes the paper.

A preliminary version of this paper has appeared in [10]. Novel contributions of the present paper with respect to [10] are the self-optimizing algorithm and its convergence analysis presented in III, more general traffic models, and additional numerical experiments for the learning procedure of Section V. The efficiency of local optima is compared with the global optimum, and the influence of correlated users arrivals is analyzed.

II. Dimensioning

A. System model

We consider the downlink scenario of a wireless network where users arrive at random times and locations, to receive a file of random size σ, with $\mathbb{E}[\sigma] < +\infty$. We assume that there is no user mobility and that users leave the network upon service completion. We denote by $\mathbb{A} \subset \mathbb{R}^2$ the network area which we assume to be bounded. $\mathbb{A}$ contains a BS (alternatively denoted as macro-cell) and several RSs. We denote by N_R the number of RSs, and we use the convention that station 0 is the BS and station s , $1 \leq s \leq N_R$ is the s-th RS.

We use the terminology of point processes to state assumptions on the arrival process clearly. We denote by $\{T_k, r_k, \sigma_k\}_{k \in \mathbb{Z}}$ the users' instants of arrival, their location and their file size. For $B \subset \mathbb{R} \times \mathbb{A}$ a Borel set, we define the number of users who arrive in B:

$$N(B) = \sum_{k \in \mathbb{Z}} \mathbf{1}_B(T_k, r_k), \tag{1}$$

and the measure of the arrival process m:

$$m(B) = \mathbb{E}\left[N(B)\right]. \tag{2}$$

We define $\mathcal{F}_t$ the σ-algebra generated by:

$$\left(N(B) : B \subset (-\infty, t] \times \mathbb{A} \text{ Borel set}\right), \tag{3}$$

which represents the available information when observing the arrival process up to time t. To ease the notation, we define $\xi_t \in \Xi$ the "effective memory" of the arrival process, with Ξ a compact metric space, so that $\mathbb{E}\left[.|\mathcal{F}_t\right] = \mathbb{E}\left[.|\xi_t\right]$. Finally, we define the conditional intensity measure of the arrival process at time t by:

$$m(B|\xi_t) = \mathbb{E}\left[N(B)|\xi_t\right]. \tag{4}$$

We will use three sets of assumptions for the arrival process:

Assumptions 1 (stationary ergodic traffic)**.** *The arrival process satisfies:*

- *Time-stationary: for* $t \in \mathbb{R}$, $\{T_k - t, r_k, \sigma_k\}_{k\in\mathbb{Z}} = \{T_k, r_k, \sigma_k\}_{k\in\mathbb{Z}}$ *in distribution*
- *Independence between arrivals and file sizes* $\{T_k, r_k\}_{k\in\mathbb{Z}} \perp\!\!\!\perp \{\sigma_k\}_{k\in\mathbb{Z}}$
- *Ergodicity: the transformation* $\{T_k, r_k, \sigma_k\}_{k\in\mathbb{Z}} \mapsto \{T_k - t, r_k, \sigma_k\}_{k\in\mathbb{Z}}$ *is ergodic*
- *Continuity with respect to Lebesgue measure in space:* $m(dr \times dt) = \lambda(r)dr \times dt$.
- *Bounded intensity:* $\sup\limits_{r\in\mathbb{A}} \lambda(r) < +\infty$

Assumptions 2 (stationary ergodic light traffic)**.** *The arrival process satisfies assumptions 1 and:*

- *Light arrivals: for* $T \geq 0$, $\mathbb{E}\left[N([0,T] \times \mathbb{A})^2\right] < +\infty$
- *Conditional continuity with respect to Lebesgue measure in space:* $\exists\overline{\lambda}$, $m(dr \times [0,T)|\xi_0) = \overline{\lambda}(r, [0,T), \xi_0)dr$.
- *Bounded conditional intensity:* $\sup\limits_{\xi\in\Xi} \sup\limits_{r\in\mathbb{A}} \overline{\lambda}(r, [0,T), \xi_0) < +\infty$

Assumptions 3 (Poisson light traffic)**.** *The arrival process satisfies assumptions 2 and is a Poisson process:*

- $N(B)$ *is a Poisson random variable with mean* $m(B)$
- $(N(B_1), \ldots, N(B_N))$ *are independent if* $\cap_{n=1}^{N} B_n = \emptyset$.

It is noted that assumptions 1 are the most general, allowing for correlated arrivals in both time and space, while 3 is the most restrictive. A special case of assumptions 2 is Markov modulated Poisson arrivals: $t \to \xi_t$ is a Markov process whose evolution is independent of the arrival process, and given $\{\xi_t\}_t$, the arrival process is a Poisson process. It is also noted that we do not assume that σ has finite variance so that our results hold for heavy-tailed traffic.

As mentioned earlier, RSs have no direct link to the backhaul, and are connected to the BS by a wireless link. This wireless link uses the same radio resources as the station to users' links and we are interested in finding an appropriate resource sharing method. This mechanism is often called in-band relaying. Depending on the multi-access radio technology, the radio resources can refer to codes in Code Division Multiple Access (CDMA), to time slots in Time Division Multiple Access (TDMA) or to time-frequency blocks in Orthogonal Frequency-Division Multiple Access (OFDMA). We ignore the granularity of resources and we denote by $x \in [0,1]$ the proportion of resources allocated to the link between the BS and RSs. We further assume that Round Robin (RR) scheduling applies in all links: the link between the BS and RSs is shared in a Processor Sharing (PS) way among the RSs, and that each link between a station and the users it serves is shared in a PS way among those users.

B. System capacity

Let $\mathbb{A}_s \subset \mathbb{A}$ denote the area covered by station s. We denote by μ the Lebesgue measure. For a given $x \in [0,1]$ we now calculate the capacity of the system, and the optimal resource sharing strategy x^* which ensures stability whenever it is possible. We assume until the end of this section that the traffic is uniform $m(dr \times dt) = \lambda_0 dr \times dt$. Namely, we denote by C the capacity of the system defined as the maximal value of $\lambda_0\mathbb{E}[\sigma]$ that keeps the system stable i.e the number of users in the system does not grow to infinity. We write $R_{rel,s}$, $1 \leq s \leq N_R$ the data rate of the link between BS and RS s when it is the only active link, and $R_s(r)$, $r \in \mathbb{A}_s$ the data rate between station s and a user located at r when he is alone in the system. The effect of inter-cell interference is incorporated in $R_{rel,s}$ and $R_s(r)$, hence the results given here hold regardless of the amount of inter-cell interference.

Theorem 1. *The capacity C of the system is:*

$$C(x) = \min\left(C_{rel}(x), \min_{0\leq s\leq N_R} C_s(x)\right), \tag{5}$$

with:

$$C_{rel}(x) = x\left(\sum_{s=1}^{N_R}\frac{\mu(\mathbb{A}_s)}{R_{rel,s}}\right)^{-1}, \tag{6}$$

$$C_s(x) = (1-x)\left(\int_{\mathbb{A}_s}\frac{1}{R_s(r)}dr\right)^{-1}. \tag{7}$$

Furthermore, there exists a unique $x^* \in [0,1]$ which maximizes the capacity,

$$x^* = \frac{\left(\max\limits_{0\leq s\leq N_R}\int_{\mathbb{A}_s}\frac{1}{R_s(r)}dr\right)^{-1}}{\left(\max\limits_{0\leq s\leq N_R}\int_{\mathbb{A}_s}\frac{1}{R_s(r)}dr\right)^{-1} + \left(\sum_{s=1}^{N_R}\frac{\mu(\mathbb{A}_s)}{R_{rel,s}}\right)^{-1}}, \tag{8}$$

with $C^* = C_{rel}(x^*) = C_s(x^*)$ the maximal capacity.

Proof: See appendix A. ■

It is noted that this result applies regardless of the underlying packet dynamics. More precisely, consider two scenarios:

1) Small files: When a user served by a RS arrives in the network, the file he wants to receive enters the BS to RSs link and once the whole file has gone through that link, it enters the corresponding RS to user link and is transmitted. This model is reasonable for small files.
2) Larger Files: In a more realistic setting, when a user served by a RS arrives in the network, the file he wants to receive arrives as small packets which enter the BS to RSs link, possibly with delays between packets. Once a packet has gone through the BS to RSs link it immediately enters the RS to user link. Here the file can be "split" between the two successive links.

For both traffic models the demonstration remains the same, and the system capacity does not change.

C. Relay gain

We now introduce the concept of RS placement gain, and give a method to evaluate the resulting capacity improvement. We assume that the signal attenuation per distance unit is smaller for the useful signal between the BS and RSs than for interfering signals. This can be achieved by placing RSs high enough so that the propagation between the BS and RSs is close to the line-of-sight case, while taking advantage of buildings to increase the attenuation of interfering signals. Assume that the propagation loss at distance $\|r\|$ is $\frac{A}{\|r\|^{\eta_r}}$ with $2 \leq \eta_r \leq \eta$ for the useful signal between the BS and RSs, and $\frac{A}{\|r\|^{\eta}}$ for all other signals. The case $\eta_r = 2$ corresponds to line-of-sight propagation between BS and RSs. We call $\eta - \eta_r$ the relaying gain, and $\eta_r = 2$ gives an upper bound on the achievable capacity by intelligent relay placement.

D. Numerical experiments

We now evaluate the influence of the system parameters on the performance using a classical model. The model parameters are given in Table I, and Figure 1 represents the network layout. Interference from neighbouring cells is taken into account. We now state the ergodic throughput $R_s(r)$ calculation method in the OFDMA case. Assuming that the fast-fading is a multiplicative random variable of mean 1, we have that:

$$R_s(r) = N_{RB} \int_{\mathbb{R}^+} \phi(SINR_s(r)y)p(y)dy, \tag{9}$$

with N_{RB} the number of resource blocks, ϕ - a link-level curve mapping instantaneous Signal to Interference plus Noise Ratio (SINR) into data rate on a resource block, $SINR_s(r)$ - the mean SINR at $r \in \mathbb{A}_s$ and p the probability density function (p.d.f) of the fast-fading. In the Rayleigh case, $p(y) = e^{-y}$. Similar models apply in the TDMA and CDMA case (see for example [11], [12]). It is noted that we choose a large cell radius since [13] had shown that relays are only beneficial in such a setting.

Model parameters	
Cell layout	Hexagonal
Antenna type	Omnidirectional
Cell Radius	$2km$
Access technology	OFDMA
Fast-fading model	Rayleigh
N_{RB}	10
Resource block size	$180kHz$
BS transmit power	$46dBm$
RS maximum transmit power	$30dBm$
Thermal noise	$-174dBm/Hz$
Path loss model	$128 + 37.6\log_{10}(d)$ dB, d in km
File size	$10Mbytes$

TABLE I
MODEL PARAMETERS

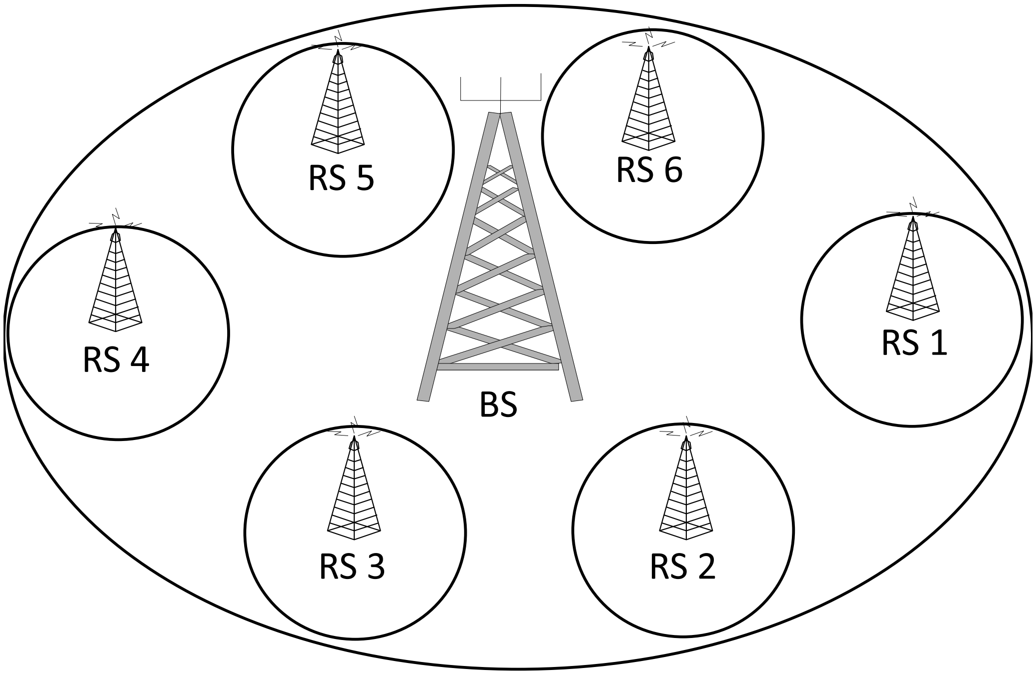

Fig. 1. Relay placement

Figure 2 and 3 show the capacity of the system and the optimal relay transmit power respectively as the number of relays grows, with and without relaying gain. The optimal relay transmit powers are determined through exhaustive

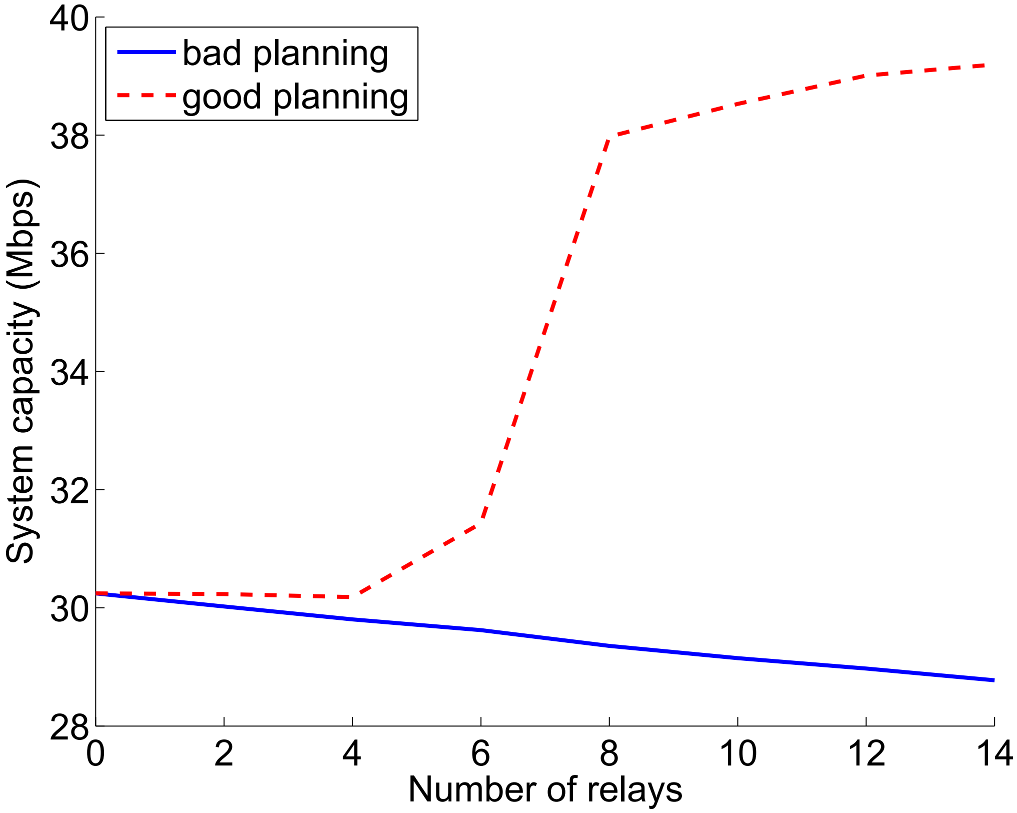

Fig. 2. System capacity as a function of the number of relays, for different planning strategies

search for a discrete set of possible values ($\{-10, \ldots, 60\}$ dBm), all relays having the same transmit power. The case without relaying gain is denoted "bad planning" and with relaying gain "good planning". It is noted that the value of the optimal relay transmit power in the "bad planning" case is $0mW$ for all number of relays (below the x-axis). It demonstrates that the impact of relaying gain is fundamental since without relaying gain it is actually detrimental to deploy relays. With relaying gain however, the system capacity increases sharply.

Figure 4 shows the impact of the relaying gain on the system capacity for a fixed number of relays (15 in this case), and we can see that the capacity increases almost linearly in the relaying gain. This can be explained by the fact that $\log_2(1 + S\|r\|^{\eta-\eta_r})$ is close to $\log_2(S) + (\eta - \eta_r)\log_2(\|r\|)$ when $S\|r\|^{\eta-\eta_r}$ is large. It shows that if one is able to evaluate the relaying gain prior to deployment (by measuring the value of the path loss exponent in candidate sites for relay placement), one can actually determine if relay deployment is beneficial and the expected benefit. Furthermore the point where the two curves intersect represents the minimal relaying gain needed for any benefit from relay deployment to appear.

III. Self-Optimization

We have given a procedure for network dimensioning and we now show that the network can adapt itself to traffic variations based solely on measurements and perform automatic traffic balancing. Two critical parameters are tuned: the pilot powers of the RSs which control the zone served by the RSs and the resources allocated to the backhaul links. Both parameters are updated simultaneously, and we show that the mechanism proposed ensures their coordination. Previous work in [14] used a similar approach to tune the transmitted pilot powers of BSs. We show here that, in relay enhanced networks, we can tune the transmitted pilot powers and the resource allocation to the backhaul and converge to an optimal configuration. Unlike the previous section, we consider a slightly more general model: the resources allocated to the backhaul links are not shared in a PS manner any more.

Instead of sharing $\sum_{s=1}^{N_R} x_s$ resources among the backhaul links in a PS manner, for each s, a quantity x_s is allocated to the link between the BS and RS s, which does not require a scheduler to share the resources among

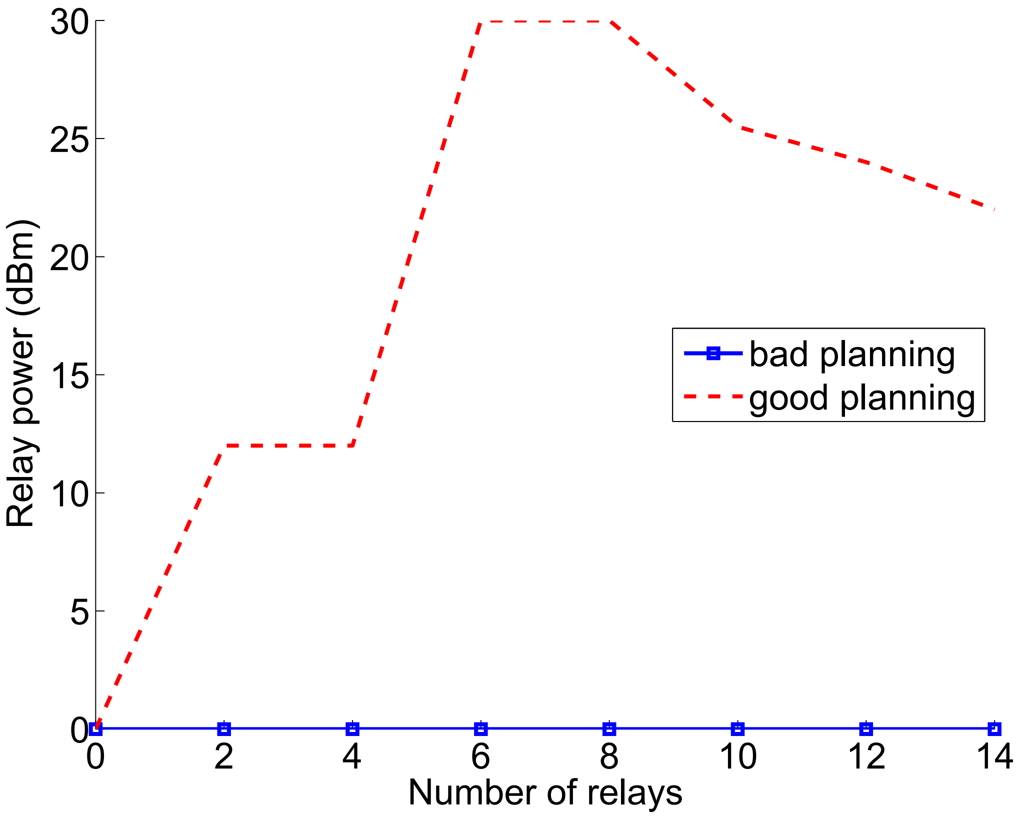

Fig. 3. Optimal relay transmit power as a function of the number of relays, for different planning strategies

the different backhaul links. If PS applies for the backhaul links then, the quantity allocated to the backhaul is simply $\sum_{s=1}^{N_R} x_s$.

A. Traffic estimation

In appendix B, we show that quantities of interest can be estimated by traffic measurements. We do not assume the traffic to be uniform. We write $\overline{\rho_s}$ the load of station s and $\overline{\rho}_{rel,s}$ the load of the backhaul between the BS and RS s, which can be expressed as:

$$\overline{\rho}_s = \frac{\mathbb{E}\left[\sigma\right]}{1-\sum_{s'=1}^{N_R} x_{s'}} \int_{\mathbb{A}_s} \frac{\lambda(r)}{R_s(r)} dr \; , \; \overline{\rho}_{rel,s} = \frac{\mathbb{E}\left[\sigma\right] \int_{\mathbb{A}_s} \lambda(r) dr}{x_s R_{rel,s}}. \tag{10}$$

Define ρ_s and $\rho_{rel,s}$ by :

$$\rho_s = \int_{\mathbb{A}_s} \frac{\lambda(r)}{R_s(r)} dr \; , \; \rho_{rel,s} = \frac{\int_{\mathbb{A}_s} \lambda(r) dr}{R_{rel,s}}. \tag{11}$$

then the loads can be expressed in the reduced form:

$$\overline{\rho}_s = \frac{\mathbb{E}\left[\sigma\right] \rho_s}{1-\sum_{s'=1}^{N_R} x_{s'}} \; , \; \overline{\rho}_{rel,s} = \frac{\mathbb{E}\left[\sigma\right] \rho_{rel,s}}{x_s}. \tag{12}$$

The condition for load balancing is $\overline{\rho}_{rel,s} = \overline{\rho}_s = \overline{\rho}_0$, which reduces to:

$$\frac{\rho_{rel,s}}{x_s} = \frac{\rho_s}{1-\sum_{s'=1}^{N_R} x_{s'}} = \frac{\rho_0}{1-\sum_{s'=1}^{N_R} x_{s'}}. \tag{13}$$

The mean flow size $\mathbb{E}\left[\sigma\right]$ has disappeared, so that load balancing can be achieved without estimating it.

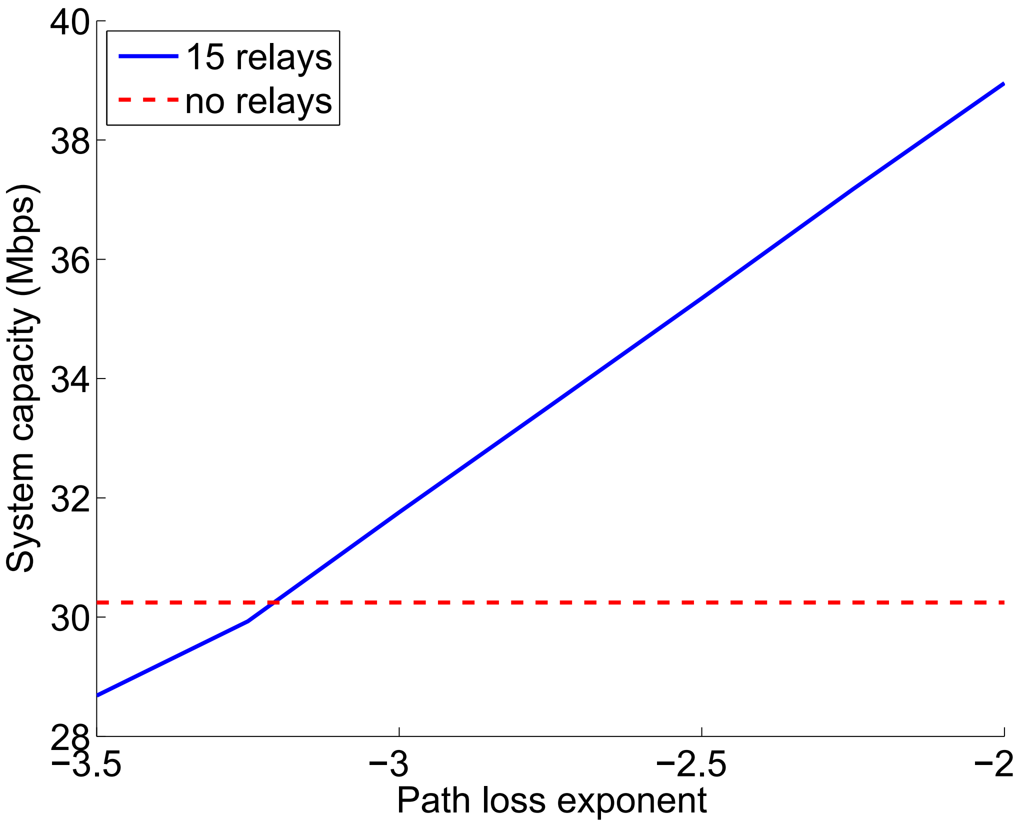

Fig. 4. Impact of the relaying gain on the system capacity

Time is slotted, with T the time slot size. The n-th time slot is $[nT,(n+1)T)$. We write $\xi[n] = \xi_{Tn}$. According to theorem 4, the loads can be estimated by:

$$\rho_s[n] = \frac{1}{T}\sum_{k\in\mathbb{Z}} \frac{1}{R_s(r_k)} \mathbf{1}_{\mathbb{A}_s}(r_k)\mathbf{1}_{[nT,(n+1)T)}(T_k), \tag{14}$$

$$\rho_{rel,s}[n] = \frac{1}{T}\sum_{k\in\mathbb{Z}} \frac{1}{R_{rel,s}} \mathbf{1}_{\mathbb{A}_s}(r_k)\mathbf{1}_{[nT,(n+1)T)}(T_k) \tag{15}$$

Assumptions 4. *(i)* $\inf_{r\in\mathbb{A}} \min_s R_s(r) = R_{min} > 0$
(ii) $P \to \mu(\mathbb{A}_s(P))$ *is Lipschitz continuous on* $\mathcal{P} = [P_{min}, P_{max}]^{N_R+1}$ *with* $0 < P_{min} \le P_{max} < +\infty$.

(i) is valid as long as there is an admission control rule on the minimal data rate for a user to enter the system. Conditions for (ii) to hold were given in [14]. And they imply that $P \to \rho_{rel,s}(P)$ and $P \to \rho_{rel,s}(P)$ are Lipschitz continuous. For the classical model where signal attenuation is taken as $\frac{A}{d^\eta}$ with d the distance between transmitter and receiver and A, η two positive constants, the assumption is valid.

Theorem 4 states that the load estimates are unbiased:

$$\mathbb{E}\left[\rho_s[n]\right] = \rho_s \ , \ \mathbb{E}\left[\rho_{rel,s}[n]\right] = \rho_{rel,s}. \tag{16}$$

B. Traffic balancing for the backhaul

First assume that the RSs transmit powers are fixed, so that the zones they serve do not change. We want to balance the traffic based on measurements, starting from an arbitrary allocation. If $\mathbb{A}_s$ has Lebesgue measure 0 we can simply ignore RS s, hence we will assume, without loss of generality, that $\min_s \rho_s > 0$ and $\min_s \rho_{rel,s} > 0$.

Proposition 1. *(i) The unique solution* (13) *is $x^*(\rho)$:*

$$x^*_s(\rho) = \frac{\frac{\rho_{rel,s}}{\rho_s}}{1 + \sum_{s'=1}^{N_R} \frac{\rho_{rel,s'}}{\rho_{s'}}} \tag{17}$$

(ii) We have that $0 < \sum_{s'=1}^{N_R} x_{s'}(\rho) < 1$
(iii) $\rho \to x^*(\rho)$ is locally Lipschitz continuous

Proof: (i) is proven by noticing that for any solution we must have that

$$s \to \frac{x_s \rho_s}{\rho_{rel,s}} = 1 - \sum_{s'=1}^{N_R} x_{s'} \tag{18}$$

is constant. (ii) is straightforward since $0 < \frac{\rho_{rel,s}}{\rho_s} < +\infty$ and equation (17). (iii) Is true since we have assumed $\rho_s > 0$. ■

Write $x_s[n]$ the proportion of resources allocated to the link between the BS and RS s during the n-th time slot, and $\epsilon_n > 0$ a step size. We consider two types of steps sizes:

- (constant step sizes) $\epsilon_n = \epsilon > 0$
- (decreasing step sizes) $\epsilon_n = \frac{1}{n^\gamma}$ with $\gamma_0 < \gamma \leq 1$.

We define H the admissible set which is convex:

$$H = \{x : x_s \geq 0 \ , \ 0 \leq \sum_{s=1}^{N_R} x_s \leq 1\}. \tag{19}$$

We write $[.]_H^+$ the projection on H. We consider the following iterative scheme for traffic balance:

$$x_s[n+1] = [x_s[n] + \epsilon_n g_s(\rho[n], x[n])]_H^+ \ , \tag{20}$$

$$g_s(\rho, x) = \rho_{rel,s}(1 - \sum_{s=1}^{N_R} x_s) - \rho_s x_s. \tag{21}$$

The convergence to the unique optimal point is given by the following theorem. The proof is based on stochastic approximation: we associate an Ordinary Differential Equation (ODE) to the iterative scheme and study its asymptotic behaviour. We then prove that the iterates converge to attractors of the ODE. The defintion of convergence in distribution is recalled in appendix C.

Theorem 2. *With assumptions 2 and 5, the sequence $\{x[n]\}_n$ converges to $x^*(\rho)$. The convergence occurs almost surely (a.s) for decreasing step sizes, and in distribution for constant step sizes with $\epsilon \to 0^+$.*

Proof: See appendix D. ■

C. Coordination between backhaul and cell sizes

We now assume that both the resource allocation to the backhaul, and the zones served by the relays are adapted simultaneously, and we propose a coordination mechanism. The idea is to make the two mechanisms operate on a "different time scale", namely, the backhaul adaptation is sufficiently fast compared to the cell sizes so that it appears as quasi-static. Relevant two-time scales stochastic approximation results will be used to prove convergence.

We assume that users attach themselves to the station with the strongest received pilot power. Let P_s denote the power of the pilot signal transmitted by station s and $k_s(r)$ the signal attenuation between station s and location $r \in \mathbb{A}$, the zones covered by stations can be written:

$$\mathbb{A}_s(P) = \{r : s \in \arg\max_{s'} P_{s'} k_{s'}(r)\}. \tag{22}$$

We write $P_s[n]$ the power of the pilot signal transmitted by station s during the n-th time slot. Let $\delta_n > 0$ another step sizes sequence. As previously, we distinguish two cases:

- (constant step sizes) $\epsilon_n = \epsilon > 0$, $\delta_n = \delta(\epsilon) > 0$, with $\frac{\delta(\epsilon)}{\epsilon} \underset{\epsilon \to 0^+}{\to} 0$

- (decreasing step sizes) $\epsilon_n = \frac{1}{n^{\gamma_1}}$, $\delta_n = \frac{1}{n^{\gamma_2}}$, with $\gamma_0 < \gamma_1 < \gamma_2 \leq 1$

We consider the constraint set for the pilot powers $\mathcal{P} = [P_{min}, P_{max}]^{N_R+1}$ with $0 < P_{min} \leq P_{max} < +\infty$. The update equations are:

$$x_s[n+1] = [x_s[n] + \epsilon_n g_s(\rho[n], x[n])]_H^+ \tag{23}$$

$$P_s[n+1] = [P_s[n] + \delta_n h_s(\rho[n], P[n])]_{\mathcal{P}}^+ , \tag{24}$$

$$h_s(\rho, P) = P_s(\rho_0(P) - \rho_s(P)). \tag{25}$$

The convergence to a network configuration where the loads of all links are equal is given by the next result.

Theorem 3. *With assumptions 2 and 5, the sequence $\{(x[n], P[n])\}_n$ converges to a set on which the loads of all links are equal, for P_{min} sufficiently small and P_{max} sufficiently large. As in the previous theorem, the convergence occurs a.s for decreasing step sizes, and in distribution for constant step sizes with $\epsilon \to 0^+$.*

Proof: See appendix E. ■

For the proof we will need the following result from [14][Theorem 4].

Lemma 1. *Consider the ODE:*

$$\dot{P}_s = P_s[\rho_0(P) - \rho_s(P)], \tag{26}$$

under the previous assumptions, all solutions to (26) *are defined on* $\mathbb{R}^+$, *all solutions verify:*

$$0 < \inf_{t \in \mathbb{R}^+} P_s(t) \leq \sup_{t \in \mathbb{R}^+} P_s(t) < +\infty, \tag{27}$$

and $\mathcal{L} = \{P : \min_s \rho_s(P) = \max_s \rho_s(P)\}$ *is a compact Lyapunov stable attractor for* (26).

D. Numerical experiments

We now show some numerical experiments to assess the efficiency of the proposed method. We have proven mathematically that, for a given stationary traffic, the proposed algorithms converge to the optimal configuration. However, in practical situations, the traffic changes over the course of a day, with traffic peaks and periods during which the served traffic is low, for example during the night.

Our numerical experiments show that when the traffic is not stationary, the algorithm is able to adapt itself and successfully "track" the changing traffic pattern. One BS and 4 RSs are considered. To demonstrate the tracking properties, we adopt the following traffic configuration: a uniform traffic of 50 Mbps which does not change during time, and a "hot-spot" i.e a limited zone with high traffic, located next to RS 1. The hot-spot traffic varies between 0 Mbps and 30 Mbps, and the time interval between the maximal traffic and minimal traffic is 2 hours. We show that the algorithm adapts both cell sizes and backhaul resources allocation in order to handle the variation in the traffic pattern. We compare the proposed algorithm with a reference scenario in which the network parameters are static. The network parameters are the optimal static parameters for the period in which the hot-spot traffic is 10 Mbps, the second hour with the highest load. The motivation behind such a model is a scenario in which a network engineer has chosen optimal network parameters for a uniform traffic, and an unexpected traffic pattern appears for a few hours. Such traffic variations are too fast for a human operator to modify the network parameters accordingly. This situation shows what kind of gains can be expected from network equipments that can adapt themselves automatically to hourly traffic patterns.

Figure 5 illustrates the chosen network setup. Figure 6 shows the total served traffic by the network, which is the sum of the uniform traffic (50 Mbps) and the hot-spot traffic (between 0 and 30 Mbps). Figure 7 shows the evolution of the pilot power of two relay stations scaled to their total transmitted power as a function of time, when the proposed SON algorithm is used. At low traffic periods, RS 1 transmits at a high power and covers a large area. At high traffic periods RS 1 transmits at low power in order to serve a smaller area and avoid being overloaded since it absorbs most of the hot-spot traffic. Figure 8 and 9 compare the loads of links between the proposed SON algorithm and the reference scenario. In the reference scenario, the loads of the BS and of RS 1 are imbalanced, and during the high traffic periods RS 1 absorbs too much traffic, its load being close to 100%. This is highly problematic: without admission control, the average file transfer time becomes infinite when the load goes to 100%. With admission control, a load close to 100% results in unacceptably high blocking rate. With the proposed algorithm, the loads of all links are very close to each other, and are lower than in the reference scenario. At high traffic periods, the worst load is 70% which is a large improvement with respect to the reference scenario. This shows that the proposed algorithm successfully balances the loads and reduces congestion by adapting to the changing traffic pattern.

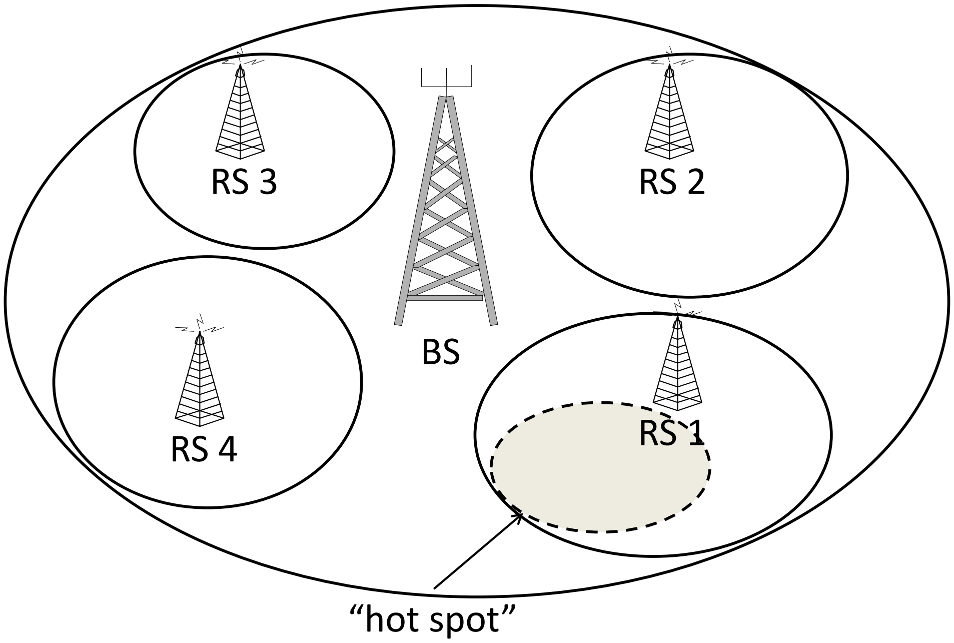

Fig. 5. Hot-spot traffic model

IV. Optimal dynamic resource allocation strategy

In the previous sections, our approach was to adapt the network to the traffic configuration, defined in terms of arrival rates. The aim was to find the best static parameters for a given traffic. We now turn to a case in which we act on a faster time scale, and instead of adapting to the arrival rates, we adapt to the current number and locations of active users. It is indeed a faster time scale since the arrival rates change on the time scale of hours, whereas the configuration of active users changes on a time scale of seconds. The BS observes the current state of the network and decides whether to activate the BS to RSs s or the stations to users' links.

A. *Infinite buffer case: stabilizing policy*

We partition each $\mathbb{A}_s$ into N regions $\mathbb{A}_{s,i}$, $1 \leq i \leq N$, each associated with a different radio condition. We call i-th traffic class in station s the users who arrive in $\mathbb{A}_{s,i}$. The state of the system can then be described by a vector $\mathbf{S} \in \mathbb{N}^{(2N_R+1)N}$, $\mathbf{S} = ((S_{s,i})_{0 \leq s \leq N_R, 1 \leq i \leq N}, (S_{rel,s,i})_{1 \leq s \leq N_R, 1 \leq i \leq N})$. In the small files framework we count the number of users present in the links, otherwise we count the number of packets. Hence $S_{s,i}$ is the number of users (packets respectively) of class i served by the station to user link in station s , and $S_{rel,s,i}$, $s \geq 1$ - the number of users (packets respectively) of class i served by the BS to RS s link. We write $R_{s,i}$ the data rate of a user of class i served by station s.

We first assume infinite buffer lengths and we want to find the policy that keeps the system stable whenever that is possible. The problem is in fact a particular case of the constrained queuing systems considered by [15]. It has

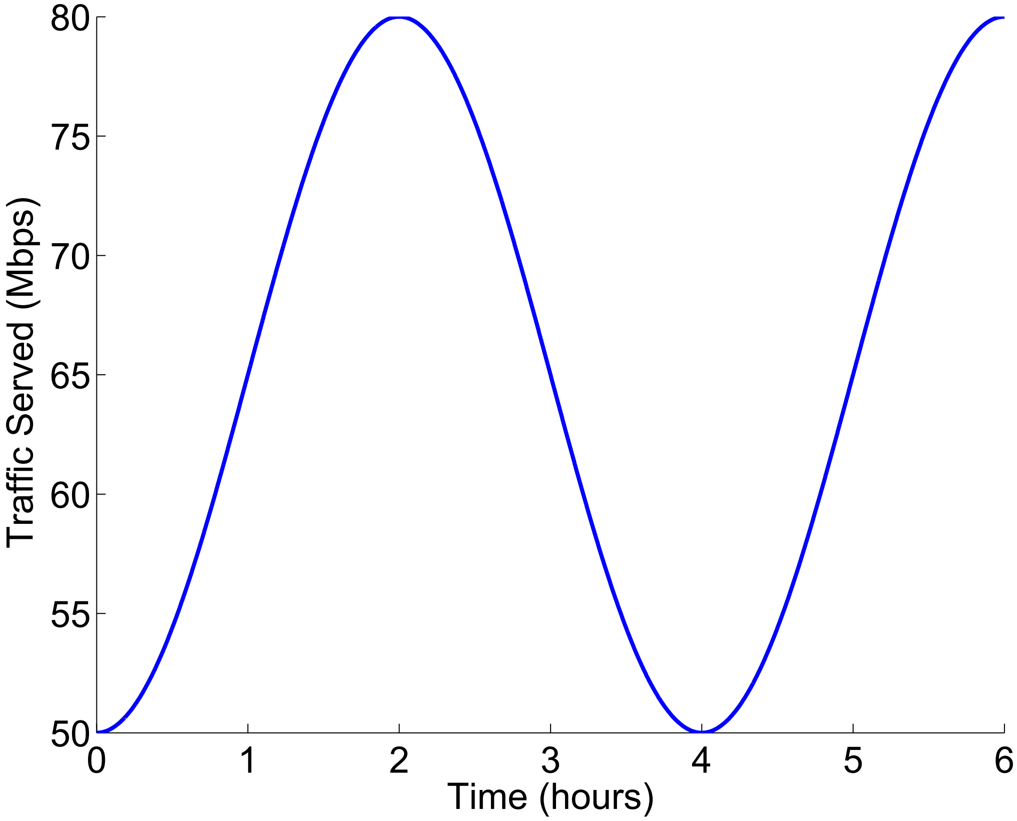

Fig. 6. Total served traffic as a function of time

been proven that such a policy exists and that it is a max-weight policy. We define the weights:

$$D_s = \max_{1 \leq i \leq N} (S_{s,i} R_{s,i}) \, , \, 0 \leq s \leq N_R \tag{28}$$

$$D_{s,rel} = \max_{1 \leq i \leq N} ((S_{rel,s,i} - S_{s,i}) R_{rel,s}) \, , \, 1 \leq s \leq N_R \tag{29}$$

The max-weight policy is then:

- *If* $\sum_{1 \leq s \leq N_R} D_{s,rel} \geq \sum_{0 \leq s \leq N_R} D_s$: activate the BS to RS s^* link with $s^* = \arg\max_{1 \leq s \leq R_S} D_{s,rel}$,
- *Else:* activate the stations to users' links, and in each station s serve the class of users $i^*_s = \arg\max_{i} n_{s,i} R_{s,i}$

B. Finite buffer case: MDP formulation

We now assume that the system state $\mathbf{S}$ is restrained to $\mathcal{S} \subset \mathbb{N}^{(2N_R+1)N}$ with $\mathcal{S}$ finite due to admission control mechanisms. We formulate the problem as a Continuous Time Markov Decision Process (CTMDP) and optimize Quality of Service (QoS) metrics such as blocking rate or file transfer time. We formulate the problem in the small files framework since we want to solve the MDP iteratively, in order to keep the state space relatively small. The learning approach of the next section however can handle large state spaces as will be demonstrated.

1) State and action spaces: We assume that each link has a maximal number of simultaneous active users.

$$\mathcal{S} = \big\{ \mathbf{S} : S_{rel,s,i} \leq \overline{S_{rel,s,i}} \, , \, 1 \leq s \leq N_R \, , \, 1 \leq i \leq N \\ \text{and } S_{s,i} \leq \overline{S_{s,i}} \, , \, 0 \leq s \leq N_R \, , \, 1 \leq i \leq N \big\}$$

We define $\mathcal{A} = \{0, 1\}$ the action space, with the convention:

- $a = 0$: activate BS to RSs links and share them in a PS sharing manner
- $a = 1$: activate stations to users' links and share them in a PS sharing manner

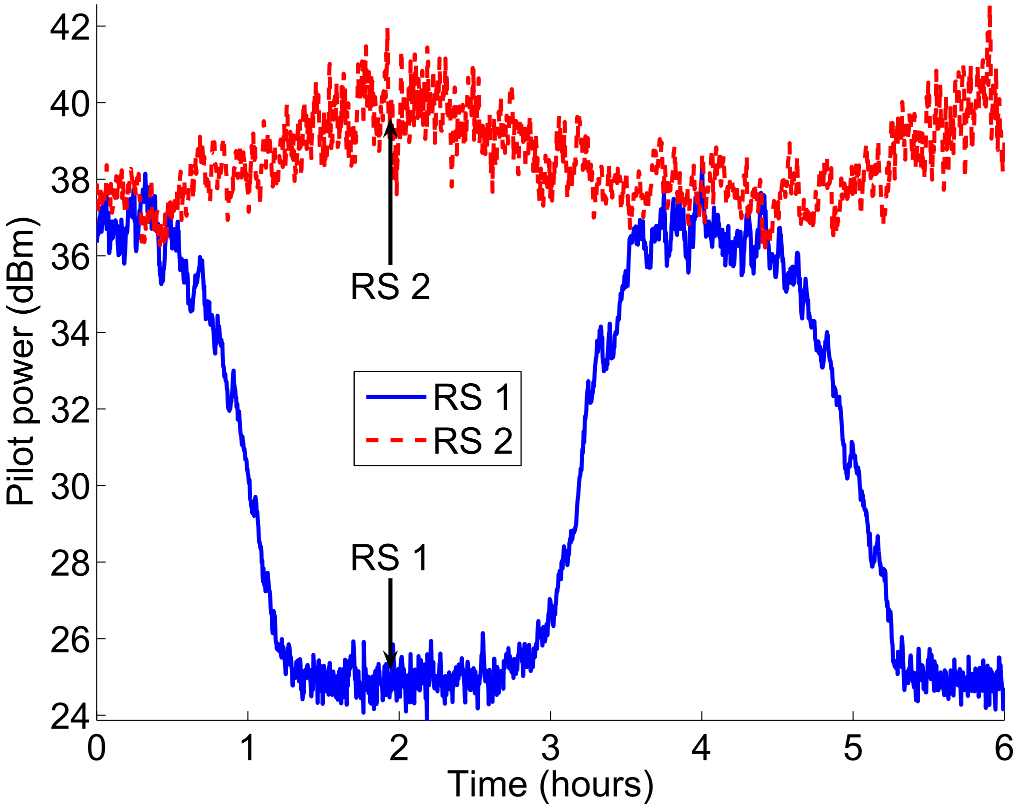

Fig. 7. SON algorithm: scaled relay pilot power as a function of time

2) Transition probabilities: Assuming that file size σ is exponentially distributed, the system is a CTMDP. Transitions from $\mathbf{S}$ to $\mathbf{S}'$ given action a have the following intensities:

- *Arrival of a user from class i in the BS:* $\mathbf{1}_{\mathcal{S}}(\mathbf{s}') \int_{\mathbb{A}_{0,i}} \lambda(r)dr$
- *Arrival of a user from class i in the BS to RS s link:* $\mathbf{1}_{\mathcal{S}}(\mathbf{s}') \int_{\mathbb{A}_{s,i}} \lambda(r)dr$
- *Departure of a user from class i in station s:* $\mathbf{1}_{\{1\}}(a)\mathbf{1}_{\mathcal{S}}(\mathbf{s}')\frac{R_{s,i}S_{s,i}}{\mathbb{E}[\sigma]\sum_{i=1}^{N} S_{s,i}}$
- *Movement of a user of class i from BS to RS s link to RS s to users' link:* $\mathbf{1}_{\{0\}}(a)\mathbf{1}_{\mathcal{S}}(\mathbf{s}')\frac{S_{rel,s,i}R_{rel,s}}{\mathbb{E}[\sigma]\sum_{i=1}^{N}\sum_{s=1}^{N_R} S_{rel,s,i}}$

3) Average reward: We call policy a mapping $\mathcal{S} \to \mathcal{D}(\mathcal{A})$, with $\mathcal{D}(\mathcal{A})$ the set of probability distributions on $\mathcal{A}$. We write $(\mathbf{S}(t), a(t), r(t))_{t\in\mathbb{R}^+}$ a sample path of the CTMDP with $\mathbf{S}(t)$ the state, $a(t)$ the action, and $r(t)$ the reward at time t respectively. We are interested in the average reward criterion of a policy P:

$$J_{\mathbf{S}_0}(P) = \lim_{T\to+\infty} \frac{1}{T}\mathbb{E}_{P,\mathbf{S}_0}\left[\int_0^T r(t)\right] \tag{30}$$

with $\mathbb{E}_{P,\mathbf{S}_0}$ the expectation with respect to the probability generated by P, starting at $\mathbf{S}_0$, which does not depend on $\mathbf{S}_0$ if the system is ergodic under policy P.

4) Performance criteria: We consider two performance criteria: mean file transfer time and blocking rate (considering admission control). For each performance criterion we can define a corresponding instantaneous reward for each state-action pair, and finding the optimal policy for the resulting MDP will yield the best policy with respect to the considered performance criterion.

To optimize the mean file transfer time, we define the reward in state $\mathbf{S}$ as the number of users divided by the arrival rate $\frac{\sum_{i=1}^{N}(S_{0,i}+\sum_{s=1}^{N_R}(S_{s,i}+S_{rel,s,i}))}{\int_{\mathbb{A}}\lambda(r)dr}$, and for any policy P that renders the system ergodic, $J_{\mathbf{S}_0}(P)$ is the mean file transfer time in the system using Little's law ([16]).

We define the blocking rate as the ratio between the mean number of blocked users and the mean number of users accessing the system, once again assuming ergodicity. Given action a, let $\beta(\mathbf{S}, a)$ the sum of transition intensities

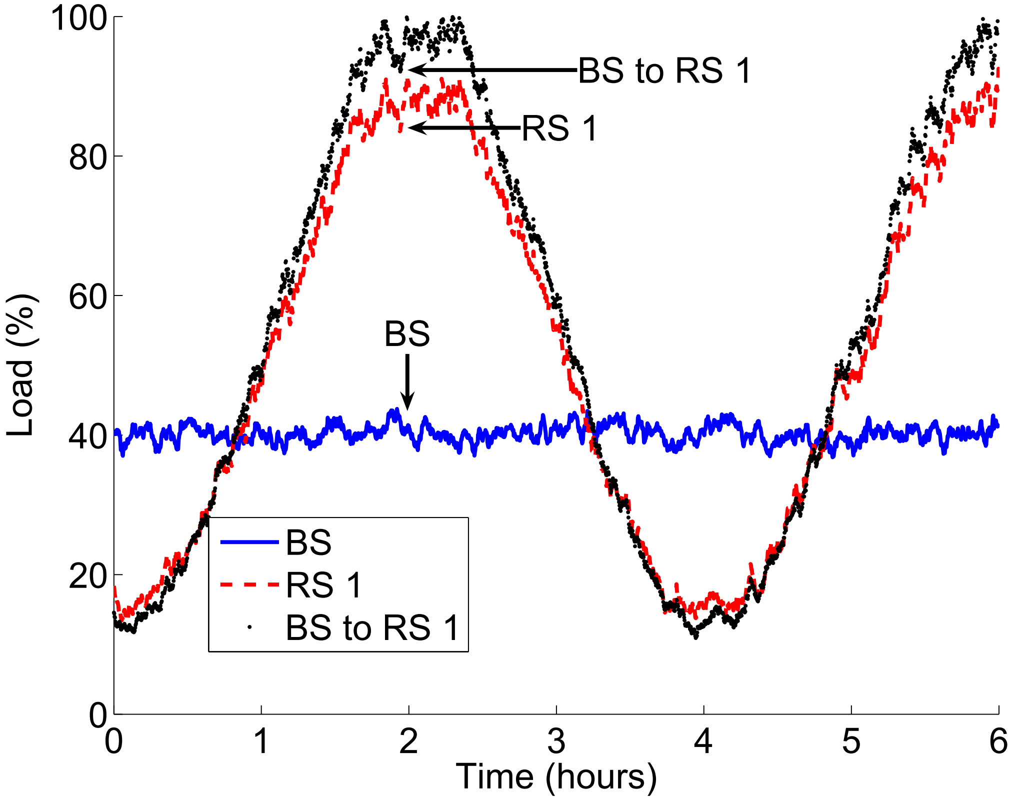

Fig. 8. Reference scenario: loads as a function of time

out of state $\mathbf{S}$ and $b(\mathbf{S}, a)$ the sum of the intensities of arrival or movements which would be blocked, then the reward is defined as $\frac{b(\mathbf{S},a)}{\beta(\mathbf{S},a)}$.

5) Optimal control and parametrization: Given the previous description, we associate a Discrete Time Markov Decision Process (DTMDP) by uniformization and we derive the optimal policy using an iterative method, by the method described in [17]. It is noted that the complexity of finding the optimal policy is exponential in the number of relays, limiting the approach to small problems. In order to preserve scalability, we introduce a well-chosen family of policies. For commodity of notation we will use the following indexing of $\mathbf{S}$: $(S_1, \cdots, S_k, \cdots, S_{(2N_R+1)N}) = ((S_{s,i})_{0 \leq s \leq N_R, 1 \leq i \leq N}, (S_{rel,s,i})_{1 \leq s \leq N_R, 1 \leq i \leq N})$. For $\theta \in \mathbb{R}^{(2N_R+1)N}$ we write

$$< \mathbf{S}, \theta >= \sum_{k=1}^{(2N_R+1)N} \theta_k S_k. \tag{31}$$

To θ we associate the deterministic weighted policy $P_{d,\theta}$:

$$P_{d,\theta}(\mathbf{S}, 1) = \begin{cases} 1 & , \quad < \mathbf{S}, \theta > \ \geq 0 \\ 0 & , \quad < \mathbf{S}, \theta > \ < 0 \end{cases} \tag{32}$$

$$P_{d,\theta}(\mathbf{S}, 0) = 1 - P_{d,\theta}(\mathbf{S}, 1) \tag{33}$$

It is noted that a deterministic weighted policy is essentially an hyperplane separating the state space in two regions, each half-space corresponding to an action of $\mathcal{A}$.

It is also noted that the max-weight policy is a deterministic weighted policy. We then compare the performance of three policies: the optimal policy, the max-weight policy and the optimal deterministic weighted policy. The optimal deterministic weighted policy is well defined since the set of deterministic policies is finite.

Figure 10 and 11 show the file transfer time and the block call rate for the three policies, for one relay, one traffic class and a maximum of 10 users for all links. We can see that the max-weight policy is very close to the optimal policy when we are concerned with the block call rate, which is natural since it attempts to ensure stability.

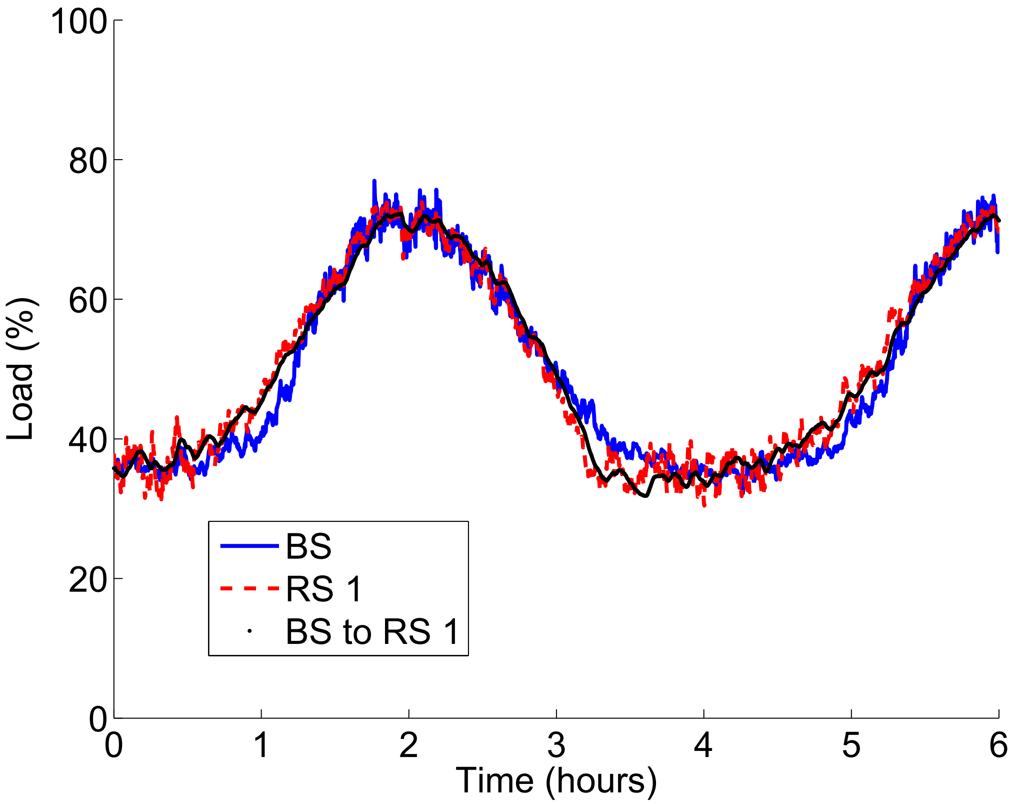

Fig. 9. SON algorithm: loads as a function of time

In the file transfer time case however, the optimal deterministic weighted policy is noticeably closer to the optimal policy than the max-weight. The fact the max-weight scheduling possibly incurs long delays has been reported in the literature. Hence based on those two results we can conclude that the set of deterministic weighted policies is rich enough to restrain the search to this set, since with a high number of relays and/or traffic classes, finding the optimal policy becomes prohibitively expensive.

V. Learning

We have demonstrated that the set of weighted policies is rich enough to represent a good trade-off between performance and search complexity. We now move on to a model-free approach, and we assume no knowledge of the transition intensities and rewards. We are interested in learning the best weighted policy, simply by observing sample paths of the Partially Observable Markov Decision Process (POMDP) $(\mathbf{S}(t), a(t), r(t))_{t\in\mathbb{N}}$. The model can be partially observed for various reasons. For example if user arrivals are correlated in time, the evolution of the system after t depends on the user arrivals before t, and this information is not present in $\mathbf{S}(t)$. The method presented here is valid without assuming Poisson arrivals or exponentially distributed file sizes.

A. Policy gradient approach

We use the approach introduced by [7] and extended to the average cost criterion in [8], [9]. It is noted that such algorithms work with stochastic policies, for the cost to be differentiable with respect to the policy parameter. We introduce stochastic weighted policy $P_{s,\theta}$:

$$P_{s,\theta}(\mathbf{S},0) = 1 - f(< \mathbf{S},\theta >)\, , \, P_{s,\theta}(\mathbf{S},1) = f(< \mathbf{S},\theta >) \tag{34}$$

with $f(x) = \frac{1}{1+e^{-x}}$. We are interested in finding the θ which minimizes the average cost $J_{\mathbf{S}_0}(P_{s,\theta})$. The link with the policies introduced in the previous section is that any deterministic weighted policy $P_{d,\theta}$ can be approximated arbitrarily well by a stochastic weighted policy $P_{s,K\frac{\theta}{\|\theta\|}}$, with $K \in \mathbb{R}^+$ arbitrarily large.

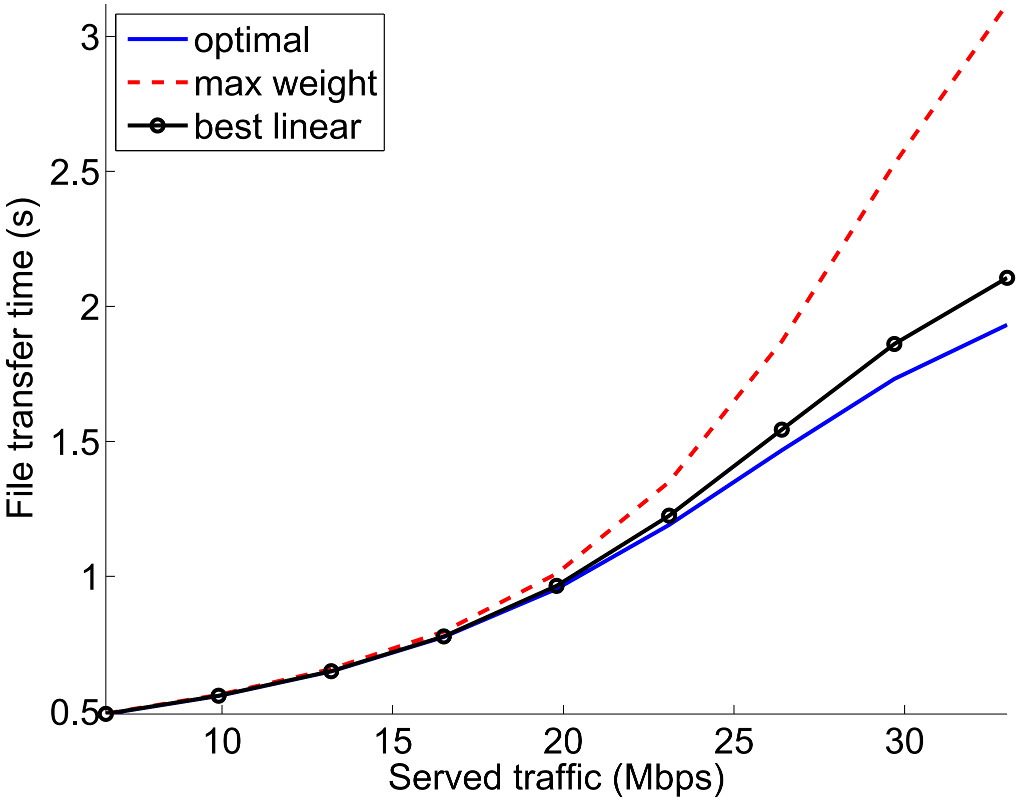

Fig. 10. File transfer time as a function of the traffic for different control strategies

B. Convergence to a local optimum

We now show how to converge to a local optimum of the average cost. We differentiate the action probabilities:

$$\frac{\partial \log(P_{s,\theta}(\mathbf{S},0))}{\partial \theta_k} = -f(<\mathbf{S},\theta>)S_k = -P_{s,\theta}(\mathbf{S},1)S_k \tag{35}$$

$$\frac{\partial \log(P_{s,\theta}(\mathbf{S},1))}{\partial \theta_k} = (1-f(<\mathbf{S},\theta>))S_k = P_{s,\theta}(\mathbf{S},0)S_k \tag{36}$$

All stochastic policies guarantee ergodicity of the system if we are considering a MDP, as stated by the next result.

Proposition 2. *If we are considering a MDP model (not a POMDP), for every θ, the Markov chain $\{\mathbf{S}(t)\}$ generated by policy $P_{s,\theta}$ is ergodic, implying that $J_{\mathbf{S}_0}(P_{s,\theta})$ is well-defined and does not depend on $\mathbf{S}_0$.*

Proof: Consider an arbitrary state $\mathbf{S}$ and the state $\mathbf{0}$. There exists a path with strictly positive probability between $\mathbf{0}$ and $\mathbf{S}$ since arrivals do not depend on the actions. There exists a path of strictly positive probability between $\mathbf{S}$ and $\mathbf{0}$ as well since in every state in which at least one user (packet) is present in the system, there is a transition corresponding to the departure of a user (or a packet) with strictly positive probability. It is the case because for any policy and any state there is a strictly positive probability for each action to be selected. This proves that the chain is irreducible.

Furthermore, the chain is aperiodic since there exists a transition from state $\mathbf{0}$ to itself. This transition exists because we have applied uniformization.

Since the state space is finite, and the chain is both irreducible and aperiodic, this proves ergodicity of the chain for any policy. ■

Using the fact that $0 < P_{s,\theta}(\mathbf{S},a) < 1$, $a \in \{0,1\}, \mathbf{S} \in \mathcal{S}$ we have that:

- $\max\limits_{a\in\{0,1\}} \max\limits_{\mathbf{S}\in\mathcal{S}} \left|\frac{\partial \log(P_{s,\theta}(\mathbf{S},0))}{\partial \theta_k}\right| < +\infty$, $1 \le k \le (2N_R+1)N$
- $\max\limits_{a\in\{0,1\}} \max\limits_{\mathbf{S}\in\mathcal{S}} r(\mathbf{S},a) < +\infty$, with $r(\mathbf{S},a)$ the reward given state $\mathbf{S}$ and action a

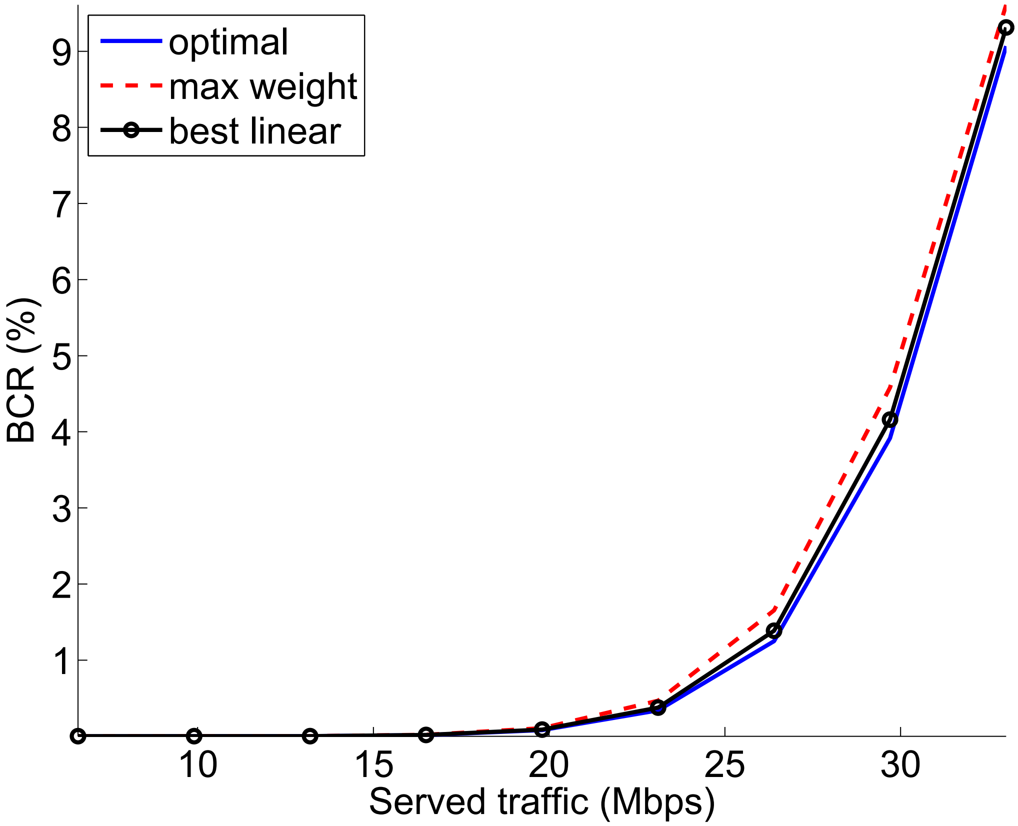

Fig. 11. Block call rate as a function of the traffic for different control strategies

Given $\beta \in (0,1)$, and a sample path of the POMDP $(\mathbf{S}(t), a(t), r(t))_{t\in\mathbb{N}}$, we define the sequence of gradient estimates and the eligibility traces $(\Delta(t), z(t))_{t\in\mathbb{N}}$ by the following recursive equation:

$$z(0) = 0 \ , \ \Delta(0) = 0 \tag{37}$$

$$z(t+1) = \beta z(t) + \nabla_\theta \log(P_{s,\theta}(\mathbf{S}(t), a(t))) \tag{38}$$

$$\Delta(t+1) = \Delta(t) + \frac{1}{t+1}[r(t)z(t) - \Delta(t)] \tag{39}$$

Furthermore [8][Theorem 4] states that: $\Delta(t) \underset{t\to+\infty}{\to} \Delta_\infty(\theta)$ almost surely and that the dot product between $\Delta_\infty(\theta)$ and $\nabla_\theta J(\theta)$ is positive. In other words, for a given θ, the limit of $-\Delta(t)$ is a descent direction. We consider $\Theta \subset \mathbb{R}^{(2N_R+1)N}$ a compact and convex set, $[.]^+_\Theta$ the projection on Θ, $(\epsilon_n)_{n\in\mathbb{N}}$ a sequence of positive step sizes (satisfying the Wolfe conditions) and we define θ_n by:

$$\theta_0 \in \Theta \tag{40}$$

$$\theta_{n+1} = [\theta_n - \epsilon_n \Delta_\infty(\theta)]^+_\Theta \tag{41}$$

then we have that $\theta_n \underset{n\to+\infty}{\to} \theta_\infty$ with θ_∞ a local minimum of J in Θ by a simple descent argument. θ_∞ is not necessarily unique if J or Θ are not convex.

Furthermore, since $-\Delta_\infty(\theta)$ is a descent direction, we have that the performance of the system improves monotonically, which is a very interesting property for system implementation. This is in sharp contrast with the traditional “learning phase” of learning algorithms such as Q-learning ([18]) when the average reward changes rapidly.

The learning method converges to a locally optimum policy if $\{\theta_n\}_n$ converges to θ_∞ a local optimum of the cost. It is noted that convergence of the controller parameter θ implies convergence of policies.

C. *Implementation issues: assumptions on traffic and scalability*

It is noted that the learning method is valid regardless of the statistical assumptions on traffic. Namely the validity of the policy gradient approach was shown by [8] even in the partially observable case.

It is noted that the algorithm is fully scalable (linear complexity) when the number of relays increase since all the components of the descent direction $\Delta_{\infty}(\theta)$ are estimated from the same sample path of the POMDP, incurring no additional costs when N_R or N increases. This is fundamental since some deployment scenarios include 30 RSs per BS.

D. *Numerical experiments*

We now evaluate the performance of the learning algorithm in the same setting as Section IV. Figures 12 and 13 represent the evolution of the mean file transfer time and the controller parameters $(\theta_1, \theta_2, \theta_3)$ respectively during the learning period. One update of θ corresponds to 10^3 iterations of the underlying POMDP. As stated above, the mean file transfer time decreases in an almost monotonic fashion. The small variations are a numerical artefact due to the fact that the average reward is calculated on a finite number of iterations of the POMDP.

We run the learning process successively a 100 times from an initial condition randomly chosen in $[-5, 5]^{(2N_R+1)N}$, and we calculate the file transfer time at the value of θ returned by the learning procedure. We calculate the global optimum by a global search (particle swarm optimization was used here). We then plot the cumulative distribution function (c.d.f) of the performance gap between the learning process and the global optimum on figure 14. In the worst case, the gap is of 25%, and the median performance gap is 11%. Hence despite its local nature and relatively low computational complexity, the learning procedure performs quite well when compared to a global search.

We compare the results between Poisson arrivals, and arrivals according to a Markov modulated Poisson process with 2 states. Both states have equal stationary probability, the average time spent in a state is 1 minute and the arrival rate in state 2 the arrival rate in state 1 multiplied by 3. In each case we estimate the gradient of the cost, calculate the sign of it's dot product with the true gradient. If it is positive then the gradient estimate is a valid ascent direction, and the accuracy of the gradient is the probability of this dot product being positive. We plot the gradient accuracy as a function of the length of the simulation on figure 15. As expected, the accuracy is less for Markov modulated arrivals than for Poisson arrivals, since the arrivals tend to be more bursty, but the gap is not very large. This suggests that the learning procedure has good numerical performance even when the arrivals are correlated.

VI. Conclusion

We have considered the problem of self-organized relays in a cellular network. The optimal static resource sharing between BS to RSs links and stations to users' links has been derived in closed form using a queuing model. The influence of key system parameters has been investigated, showing the importance of relaying gain. For non-stationary traffic, a self-organizing algorithm has been given, and its convergence has been proven using stochastic approximation techniques. Dynamic resource sharing has been considered using two approaches: stability for infinite buffers and blocking rate and file transfer time in the presence of admission control. The optimal policy has been derived using a MDP approach, which allowed us to introduce a well-chosen subset of the policy space as a form of expert knowledge. This expert knowledge has then been used in a model-free approach in which the optimal parameterized controller is found by observation and interaction with the system. Convergence to a local optimum has been demonstrated. The performance of the system improves monotonically, which is a key property for system implementation.

Appendix A
Proof of theorem 1

The process of arrivals and service requirements is $\{T_k, \mathbf{1}_{\mathbb{A}_s}(r_k)\frac{\sigma_k}{R_s(r_k)}\}_{k\in\mathbb{Z}}$ for the link between users and station s, and $\{T_k, \mathbf{1}_{\mathbb{A}_s}(r_k)\frac{\sigma_k}{R_{rel,s}}\}_{k\in\mathbb{Z}}$ for the link between the BS and RS s. Since the arrival process is stationary ergodic,

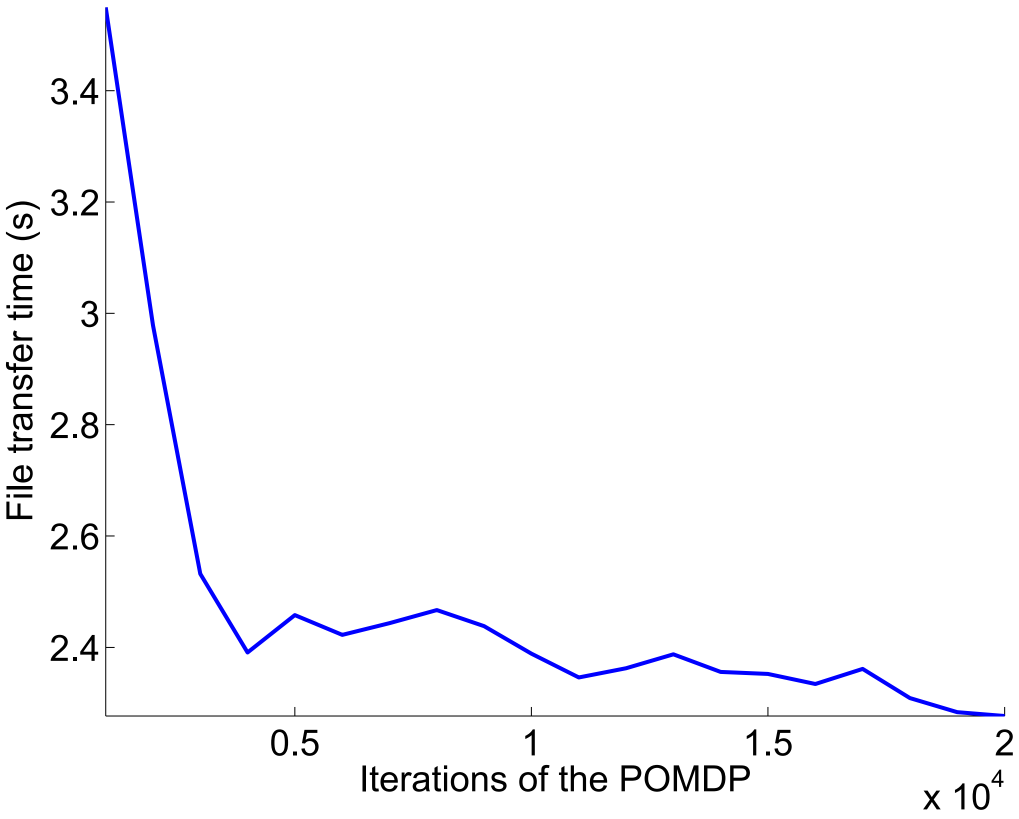

Fig. 12. File transfer time during the learning process

Loynes' theorem ([19]) gives the stability conditions:

$$\lambda_0 \mu(\mathbb{A}) \mathbb{E}\left[\sigma\right] \mathbb{E}_T^0 \left[\frac{\mathbf{1}_{\mathbb{A}_s}(r_0)}{R_s(r_0)} \right] < (1-x), \tag{42}$$

$$\lambda_0 \mu(\mathbb{A}) \mathbb{E}\left[\sigma\right] \sum_{s=1}^{N_R} \mathbb{E}_T^0 \left[\frac{\mathbf{1}_{\mathbb{A}_s}(r_0)}{R_{rel,s}} \right] < x \tag{43}$$

with $\mathbb{E}_T^0$ the Palm expectation with respect to the arrival instants. The conditions are valid for a G/G/1/PS queue since Loynes' theorem holds for all work-conserving service disciplines.

We write the capacity of the link between the BS and users:

$$C_0(x) = (1-x) \left(\int_{\mathbb{A}_0} \frac{1}{R_0(r)} dr \right)^{-1} \tag{44}$$

and the capacity of the link between the BS and RSs:

$$C_{rel}(x) = x \left(\sum_{s=1}^{N_R} \frac{\mu(\mathbb{A}_s)}{R_{rel,s}} \right)^{-1} \tag{45}$$

Now assuming that the link between the BS and RSs is stable, its output process is stationary ergodic, and using a flow conservation argument it has the same intensity as the input. The capacity of the link between RS s and its users is then:

$$C_s(x) = (1-x) \left(\int_{\mathbb{A}_s} \frac{1}{R_s(r)} dr \right)^{-1} \tag{46}$$

The stability of the system is equivalent to the stability of all queues, hence $C(x) = \min\left(C_{rel}(x), \min_{0 \le s \le N_R} (C_s(x)) \right)$. Furthermore $x \to C_{rel}(x)$ is strictly increasing and $x \to \min_{0 \le s \le N_R} (C_s(x))$ is strictly decreasing, hence the unique

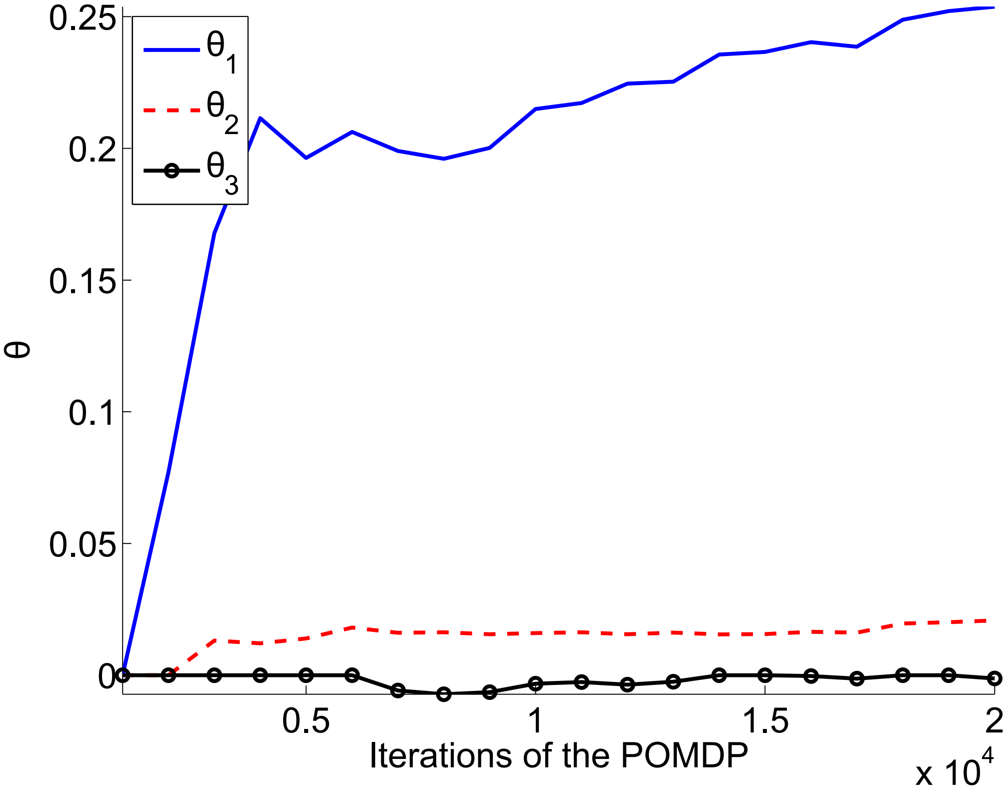

Fig. 13. Controller parameters $(\theta_1, \theta_2, \theta_3)$ during the learning process

optimal point x^* is:

$$x^* = \frac{\left(\max\limits_{0\leq s\leq N_R} \int_{\mathbb{A}_s} \frac{1}{R_s(r)} dr\right)^{-1}}{\left(\max\limits_{0\leq s\leq N_R} \int_{\mathbb{A}_s} \frac{1}{R_s(r)} dr\right)^{-1} + \left(\sum_{s=1}^{N_R} \frac{\mu(\mathbb{A}_s)}{R_{rel,s}}\right)^{-1}} \tag{47}$$

Substitution of x^* in the capacity formula yields C^* which concludes the demonstration.

Appendix B
Traffic estimation

Theorem 4. *Given $T > 0$ a measurement time, $f : \mathbb{A} \to \mathbb{R}$ - a function which is measurable, positive and bounded by $\|f\|_\infty$, we define the sequence $\{F_n\}_{n\in\mathbb{Z}}$:*

$$F_n = \frac{1}{T} \sum_{k\in\mathbb{Z}} f(r_k) \mathbf{1}_{[nT,(n+1)T)}(T_k). \tag{48}$$

We decompose F_n as a sum of its expectation, a martingale difference and a term due to the memory of the arrival process:

$$F_n = \mathbb{E}\left[F_n\right] + M_n + G_n, \tag{49}$$

$$M_n = F_n - \mathbb{E}\left[F_n|\xi_{Tn}\right], \tag{50}$$

$$G_n = \mathbb{E}\left[F_n|\xi_{Tn}\right] - \mathbb{E}\left[F_n\right]. \tag{51}$$

With assumptions 1:

$$\mathbb{E}\left[F_n\right] = \int_{\mathbb{A}} \lambda(r) f(r) dr. \tag{52}$$

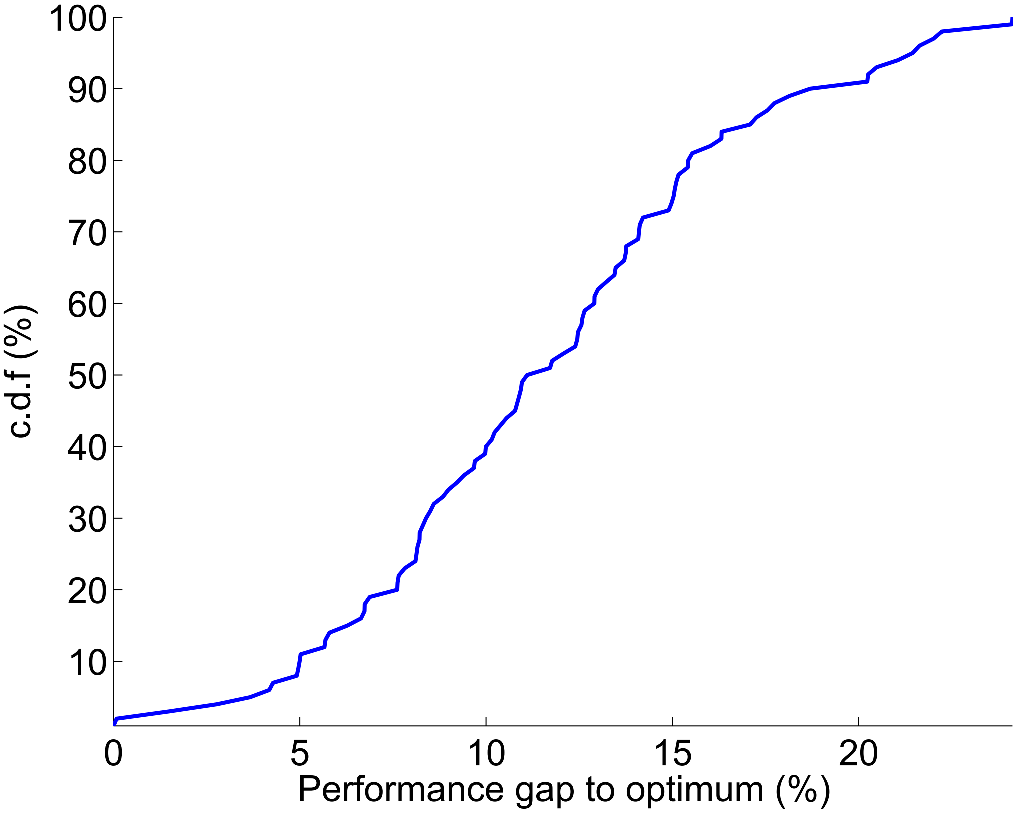

Fig. 14. Comparison between local and global optima

For assumptions 2, we further have that:

$$\sup_n \mathbb{E}\left[F_n^2\right] < +\infty, \tag{53}$$

and for $\gamma > \frac{1}{2}$:

$$\frac{1}{N^\gamma}\sum_{n=1}^{N} M_n \underset{N\to+\infty}{\to} 0\ ,\ a.s. \tag{54}$$

Furthermore:

$$\frac{1}{N}\sum_{n=1}^{N} G_n \underset{N\to+\infty}{\to} 0\ a.s. \tag{55}$$

Finally, for assumptions 3, $G_n \equiv 0$.

We introduce another assumption on the mixing properties of the arrival process which will be necessary for further results:

Assumptions 5. *There exists $\gamma_0 < 1$ such that for any measurable positive and bounded function f, if $\gamma_0 < \gamma \leq 1$:*

$$\frac{1}{N^\gamma}\sum_{n=1}^{N} G_n \underset{N\to+\infty}{\to} 0\ ,\ a.s. \tag{56}$$

It is noted that for Poisson arrivals (assumptions 3), assumptions 5 are not needed since $G_n \equiv 0$.

Proof: Applying the Campbell formula ([19]) to $(r,t) \to \frac{1}{T}f(r)\mathbf{1}_{[nT,(n+1)T)}(t)$ proves the first claim. The second claim is proven by:

$$\sup_n \mathbb{E}\left[F_n^2\right] \leq \frac{\|f\|_\infty^2}{T^2}\mathbb{E}\left[N([0,T)\times\mathbb{A})^2\right] < +\infty. \tag{57}$$

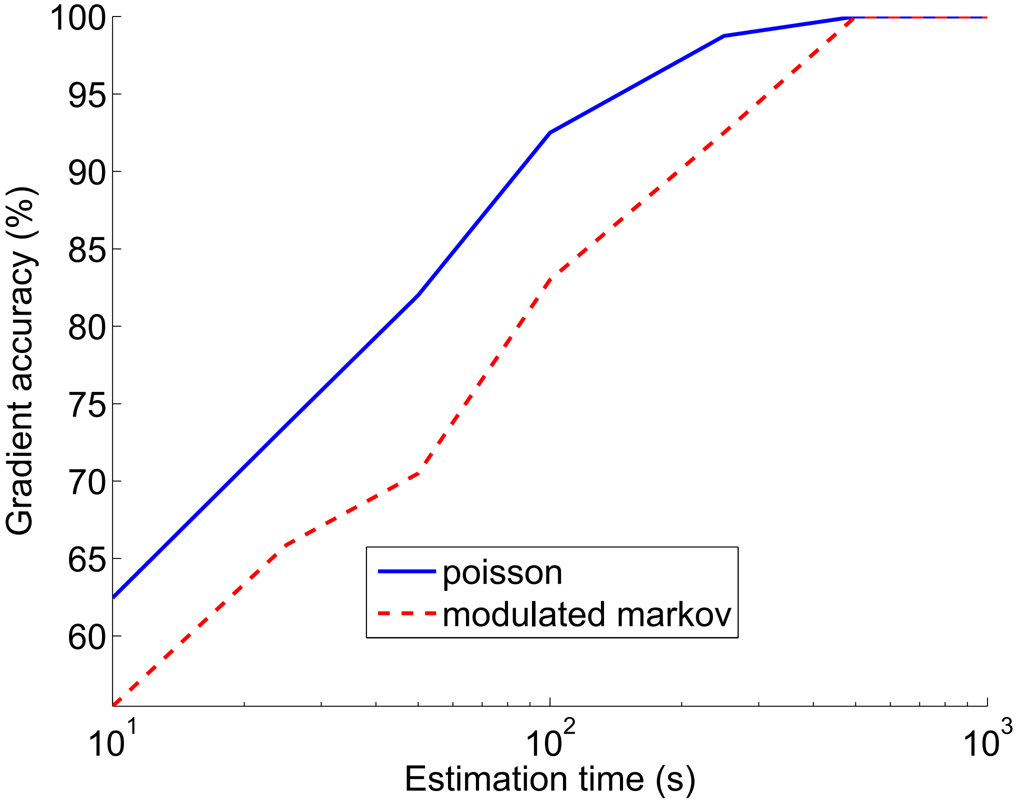

Fig. 15. Impact of correlated arrivals on gradient estimation accuracy

Define $S_N = \frac{1}{N}\sum_{n=1}^{N} M_n$. S_n is a martingale, and:

$$\mathbb{E}\left[S_n^2\right] \leq \frac{2\sup_n \mathbb{E}\left[F_n^2\right]}{n^{2\gamma-1}} \underset{n\to+\infty}{\to} 0 \text{ if } \gamma > \frac{1}{2}. \tag{58}$$

Applying the martingale convergence theorem proves the third claim. Because we have assumed ergodicity of the arrival process, $\frac{1}{N}\sum_{n=1}^{N} F_n \underset{N\to+\infty}{\to} \mathbb{E}\left[F_n\right]$ so that $\frac{1}{N}\sum_{n=1}^{N} G_n \underset{N\to+\infty}{\to} 0$ which proves the last claim. ■

Appendix C
Definition of weak convergence

We recall the notion of weak convergence, since it is used extensively in this paper, and is the correct notion for characterizing the convergence of stochastic approximation procedures when the step size is constant.

Consider iterates $\{\theta_n^{\epsilon}\}_{n,\epsilon}$ in a Euclidean space, we define the interpolated process $\theta^{\epsilon}(t)$ such that:

$$\theta^{\epsilon}(t) = \theta_n^{\epsilon} \text{ if } t = \frac{n}{\epsilon} \tag{59}$$

and $t \to \theta^{\epsilon}(t)$ is linear by parts. Let n_{ϵ}, with $\epsilon n_{\epsilon} \underset{\epsilon\to 0^+}{\to} +\infty$ We define the shifted process:

$$\overline{\theta}^{\epsilon}(t) = \theta^{\epsilon}(t + \epsilon n_{\epsilon}). \tag{60}$$

Let L a set and

$$d_L(\theta) = \inf_{\theta_L \in L} \|\theta - \theta_L\| \tag{61}$$

the distance to L. We say that the iterates $\{\theta_n^{\epsilon}\}_{n,\epsilon}$ converge weakly (or in distribution) to L if

$$d_L(\overline{\theta}^{\epsilon}(0)) \underset{\epsilon\to 0^+}{\to} 0 \text{ in probability}. \tag{62}$$

Intuitively, this means that the iterates spend most of their time near L, and the fraction of time spent near L goes to 1 when $\epsilon \to 0^+$.

Appendix D
Proof of theorem 2

ODE Since ρ does not change, we will sometimes omit it for notation clarity. Consider the ODE $\dot{x} = g(x)$, and define the Lyapunov function:

$$U(x) = \frac{1}{2}\sum_{s=1}^{N_R} \frac{g_s(x)^2}{\rho_{rel,s}}. \tag{63}$$

We calculate its gradient:

$$\frac{\partial U}{\partial x_s}(x) = \frac{\rho_s}{\rho_{rel,s}} - g_s(x) - \sum_{s'=1}^{N_R} g_{s'}(x). \tag{64}$$

Its derivative along solutions is:

$$< \nabla U, \dot{x} >= -\sum_{s=1}^{N_R} \frac{g_s(x)^2 \rho_s}{\rho_{rel,s}} - \left(\sum_{s=1}^{N_R} g_s(x)\right)^2 < 0 \tag{65}$$

It is noted that U is indeed positive definite and radially unbounded. This proves that x^* is the unique equilibrium of the ODE and that it is globally asymptotically stable. Namely all solutions of the ODE converge to x^*, regardless of the initial condition.

Projected ODE We now have to take the constraint set H into account. Namely, since the iterates are projected on H, they will follow the trajectory of the ODE *projected* on H. In the general case, we need to add a projection term to the ODE, that is:

$$\dot{x} \in g(x) + G(x). \tag{66}$$

$G(x)$ is the minimal "force" which ensures that solutions remain in the constraint set H, and $G(x) \neq \{0\}$ only if x belongs to the boundary of H. We will prove here that solutions of the (non-projected) ODE starting in H remain in it, hence $G(x) \equiv \{0\}$. If $x_s = 0$, then:

$$\dot{x_s} = \rho_{rel,s}(1 - \sum_{s=1}^{N_R} x_s) \geq 0, \tag{67}$$

and if $\sum_{s=1}^{N_R} x_s = 1$ then:

$$\frac{d}{dt}(\sum_{s=1}^{N_R} x_s) = \sum_{s=1}^{N_R} \rho_s x_s < 0. \tag{68}$$

This proves that H is an invariant set of the ODE without the need to add a projection term.

Stochastic approximation: decreasing step sizes It is noted that $x \to g(\rho, x)$ is affine hence smooth. We verify the necessary conditions for stochastic approximation theorems to be valid:

- $\sup_n \mathbb{E}\left[g_s(\rho[n], x[n])^2\right] \leq \sup_n \mathbb{E}\left[(\rho_{rel,s}[n] + \rho_s[n])^2\right] < +\infty$ from theorem 4,
- $x \to \mathbb{E}\left[g(\rho[n], x[n]) | \xi[n]\right]$ is continuous
- $\frac{1}{N^\gamma}\sum_{n=1}^{N} g(\rho[n], x[n]) - \mathbb{E}\left[g(\rho[n], x[n]) | \xi[n]\right] \underset{N \to +\infty}{\to} 0$, a.s for $\frac{1}{2} < \gamma \leq 1$ from theorem 4
- $\frac{1}{N^\gamma}\sum_{n=1}^{N} \mathbb{E}\left[g(\rho[n], x[n])\right] - \mathbb{E}\left[g(\rho[n], x[n]) | \xi[n]\right] \underset{N \to +\infty}{\to} 0$, a.s for $\gamma_0 < \gamma \leq 1$ from assumptions 5
- $x \to \mathbb{E}\left[g(\rho[n], x) | \xi[n]\right]$ is Lipschitz continuous uniformly in $\xi[n]$, because $\sup_{\xi[n] \in \Xi} \mathbb{E}\left[\rho_{rel,s}[n] + \rho_s[n] | \xi[n]\right] < +\infty$.

Applying [20][Theorem 1.1, Chapter 6, page 166] proves that the sequence $\{x[n]\}_n$ converges to $x^*(\rho)$ a.s.

Stochastic approximation: constant step sizes For constant step sizes, the following properties will be needed:

- Ξ is a compact space, and $\{\xi[n])\}_n$ does not depend on $\{x[n])\}_n$ i.e the noise process is exogenous
- $\{g(\rho[n], x[n])\}_n$ is uniformly integrable since it is bounded in mean square
- $x \to \mathbb{E}\left[g(\rho[n], x) | \xi[n]\right]$ is continuous
- $\{\mathbb{E}\left[g(\rho[n], x[n]) | \xi[n]\right])\}_n$ and $\{\mathbb{E}\left[g(\rho[n], x) | \xi[n]\right]\}_n$ are uniformly integrable since it they are bounded in mean square

- $\frac{1}{N}\sum_{n=1}^{N}(\mathbb{E}\left[g(\rho[n],x[n])\right]-\mathbb{E}\left[g(\rho[n],x[n])|\xi[n]\right]) \underset{N\to+\infty}{\to} 0$, a.s (actually the proof only requires convergence in probability)

Applying [20][Theorem 2.2, Chapter 8, page 255] proves that the sequence $\{x[n]\}_n$ converges to $x^*(\rho)$ in distribution.

Appendix E
Proof of theorem 3

According to assumption (ii), $P\to\rho(P)$ is Lipschitz continuous and all its components are bounded away from 0 on $\mathcal{P}$, hence $P\to x^*(\rho(P))$ is Lipschitz continuous as well. It is also noted that $\mathbb{E}\left[h(\rho[n],P)\right]=h(\rho(P[n]),P[n])$ by linearity.

Decreasing step sizes We have that:

- $\sup_n \mathbb{E}\left[h_s(\rho[n],P[n])^2\right]\leq \sup_n P_{max}^2\mathbb{E}\left[(\rho_s[n]+\rho_0[n])^2\right] < +\infty$ from theorem 4,
- $P\to\mathbb{E}\left[h(\rho[n],P)|\xi[n]\right]$ is continuous
- $\frac{1}{N^\gamma}\sum_{n=1}^{N}h(\rho[n],P[n])-\mathbb{E}\left[h(\rho[n],P[n])|\xi[n]\right] \underset{N\to+\infty}{\to} 0$, a.s for $\frac{1}{2}<\gamma\leq 1$ from theorem 4
- $\frac{1}{N^\gamma}\sum_{n=1}^{N}\mathbb{E}\left[h(\rho[n],P[n])\right]-\mathbb{E}\left[h(\rho[n],P[n])|\xi[n]\right] \underset{N\to+\infty}{\to} 0$, a.s for $\gamma_0<\gamma\leq 1$ from assumptions 5
- $P\to\mathbb{E}\left[h(\rho[n],P)|\xi[n]\right]$ is Lipschitz continuous uniformly in $\xi[n]$, because $P\to\mu(\mathbb{A}_s(P))$ is Lipschitz continuous, and $\sup_{\xi\in\Xi}\sup_{r\in\mathbb{A}}\overline{\lambda}(r,[0,T),\xi_0)<+\infty$.

Combining Lemma 1, and [20][Theorem 1.1, Chapter 6, page 166] for P_{min} sufficiently small and P_{max} sufficiently large , proves that the sequence $\{P[n]\}$ converges a.s to $\mathcal{L}$.

Following the same method as [21][Lemma 1, Chapter 6, page 66], we can rewrite the update equations as:

$$x_s[n+1]=\left[x_s[n]+\epsilon_n g_s(\rho[n],x[n])\right]_H^+ \tag{69}$$

$$P_s[n+1]=\left[P_s[n]+\epsilon_n\frac{\delta_n}{\epsilon_n}h_s(\rho[n],P[n])\right]_{\mathcal{P}}^+ . \tag{70}$$

In particular:

$$\left|\frac{\delta_n}{\epsilon_n}\mathbb{E}\left[h_s(\rho[n],P[n])\right]\right|\leq\frac{\delta_n}{\epsilon_n}\sup_n\sqrt{\mathbb{E}\left[h_s(\rho[n],P[n])^2\right]}$$

$$\underset{n\to+\infty}{\to} 0 \tag{71}$$

Applying [20][Theorem 1.1, Chapter 6, page 166] once again, we have that $\{x[n],P[n]\}_n$ converges a.s to the set $\{(x^*(\rho(P)),P):P\in\mathcal{P}\}$, which an asymptotically stable set for the ODE:

$$\dot{x(t)}=g(x(t)),\dot{P}(t)=0, \tag{72}$$

projected on $H\times\mathcal{P}$. Hence $\{(x[n],P[n])\}_n$ converges a.s a set on which all loads are equal for decreasing step sizes. Constant step sizes For the constant step sizes:

- Ξ is a compact space, and the noise process is exogenous
- $\{h(\rho[n],P[n])\}_n$ is uniformly integrable since it is bounded in mean square
- $P\to\mathbb{E}\left[h(\rho[n],P)|\xi[n]\right]$ is continuous
- $\{\mathbb{E}\left[h_s(\rho[n],P[n])|\xi[n]\right])\}_n$ and $\{\mathbb{E}\left[h(\rho[n],P)|\xi[n]\right]\}_n$ are uniformly integrable since it they are bounded in mean square
- $\frac{1}{N}\sum_{n=1}^{N}(\mathbb{E}\left[h(\rho[n],P[n])\right]-\mathbb{E}\left[h(\rho[n],P[n])|\xi[n]\right]) \underset{N\to+\infty}{\to} 0$, a.s (and in probability)

From Lemma 1, and [20][Theorem 2.2, Chapter 8, page 255], for P_{min} sufficiently small and P_{max} sufficiently large , this proves that $\{P[n]\}_n$ converges in distribution to $\mathcal{L}$ when $\epsilon\to 0^+$. Using the same technique as in the decreasing step size case, we write

$$x_s[n+1]=\left[x_s[n]+\epsilon g_s(\rho[n],x[n])\right]_H^+ \tag{73}$$

$$P_s[n+1]=\left[P_s[n]+\epsilon\frac{\delta(\epsilon)}{\epsilon}h_s(\rho[n],P[n])\right]_{\mathcal{P}}^+ , \tag{74}$$

and:

$$\frac{\delta(\epsilon)}{\epsilon n''}\left|\sum_{n=n'}^{n'+n''}\mathbb{E}\left[h_s(\rho[n],P[n])\right]\right| \\ \le \frac{\delta(\epsilon)}{\epsilon}\sup_n\sqrt{\mathbb{E}\left[h_s(\rho[n],P[n])^2\right]}\underset{\epsilon\to 0^+}{\to} 0 \tag{75}$$

so [20][Theorem 2.2, Chapter 8, page 255] proves that $\{x[n],P[n]\}_n$ converges in distribution to $\{(x^*(\rho(P)),P) : P\in\mathcal{P}\}$. This justifies that $\{(x[n],P[n])\}_n$ convergence in distribution to a set on which all loads are equal when $\epsilon\to 0$.

References

[1] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRAN); Overall description; Stage 2," 3GPP, TS 36.300, Sep. 2008.

[2] ——, "Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self-configuring and self-optimizing network (SON) use cases and solutions," 3GPP, TR 36.902, Sep. 2008.

[3] A. Stolyar and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse for best-effort traffic through distributed inter-cell coordination," in *IEEE INFOCOM 2009*, apr. 2009, pp. 1287 –1295.

[4] R. Combes, Z. Altman, and E. Altman, "Self-organizing fractional power control for interference coordination in OFDMA networks," in *IEEE ICC 2011*, june 2011.

[5] R. Combes, Z. Altman, M. Haddad, and E. Altman, "Self-optimizing strategies for interference coordination in OFDMA networks," in *IEEE ICC Workshops 2011*, june 2011.

[6] M. Dirani and Z. Altman, "A cooperative reinforcement learning approach for inter-cell interference coordination in OFDMA cellular networks," in *WiOpt 2010*, may. 2010, pp. 170 –176.

[7] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine Learning*, vol. 8, pp. 229–256, 1992.

[8] J. Baxter and P. L. Bartlett, "Infinite-Horizon Policy-Gradient Estimation," *Journal of Artificial Intelligence Research*, vol. 15, pp. 319–350, 2001.

[9] J. Baxter, P. L. Bartlett, and L. Weaver, "Experiments with Infinite-Horizon, Policy-Gradient Estimation," *Journal of Artificial Intelligence Research*, vol. 15, pp. 351–381, 2001.

[10] R. Combes, Z. Altman, and E. Altman, "Self-organizing relays in LTE networks: Queuing analysis and algorithms," in *CNSM 2011*, october 2011.

[11] ——, "Scheduling gain for frequency-selective rayleigh-fading channels with application to self-organizing packet scheduling," *Performance Evaluation*, Feb. 2011.

[12] R. Combes, S. Elayoubi, and Z. Altman, "Cross-layer analysis of scheduling gains: Application to LMMSE receivers in frequency-selective rayleigh-fading channels," in *WiOpt 2011*, may 2011, pp. 133 –139.

[13] L. Rong, S. Elayoubi, and O. Haddada, "Impact of relays on LTE-advanced performance," in *IEEE ICC 2010*, May 2010, pp. 1 –6.

[14] R. Combes, Z. Altman, and E. Altman, "Self-organizing load balancing in wireless networks: a flow-level perspective," in *IEEE INFOCOM 2012*, april 2012.

[15] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *Automatic Control, IEEE Transactions on*, vol. 37, no. 12, pp. 1936 –1948, Dec. 1992.

[16] J. D. C. Little, "A Proof for the Queuing Formula: L= λ W," *Operations Research*, vol. 9, no. 3, pp. 383–387, 1961.

[17] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley-Interscience, 2005.

[18] R. Sutton and A. Barto, *Reinforcement Learning, an Introduction*. MIT Press, 1998.

[19] F. Baccelli and P. Bremaud, *Elements of Queueing Theory. Palm Martingale Calculus and Stochastic Recurrences*. Springer, 2nd ed, 2003.

[20] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications 2nd edition*. Springer Stochastic Modeling and Applied Probability, 2003.

[21] V. S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.

Richard Combes was born in France in 1985. He received the Engineering Degree from ENST (Ecole Nationale Suprieure des Tlcommunications, Paris, France) in 2008 and the Master Degree in Mathematics from the university of Paris VII (France) in 2009. He is currently a Ph.D. student at Orange Labs (Issy-Les-Moulineaux, France), under the direction of Zwi Altman (Orange Labs), Eitan Altman (INRIA) and Sylvain Sorin (Paris VI). He was the recipient for the best paper award at the Conference on Network and Service Management (CNSM) in 2011. His current research interests include wireless networks, game theory and queuing theory.

Zwi Altman received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion-Israel Institute of Technology, in 1986 and 1989, and the Ph.D. degree in electronics from the INPT France in 1994. He was a Laureate of the Lavoisier scholarship of the French Foreign Ministry in 1994, and from 1994 to 1996 he was a Post-Doctoral Research Fellow in the University of Illinois at Urbana Champaign. In 1996 he joined France Telecom R&D where he has been involved in different projects related to mobile network planning, optimization and self-organization. In 2004, he co-received the France Telecom Innovation Prize, and in 2005 the IEEE Wheeler Award. From 2005 to 2007 Dr. Altman was the coordinator of the Eureka Celtic Gandalf project Monitoring and Self-Tuning of RRM Parameters in a Multi-System Network that received the Celtic Excellence Award. He is currently involved in the collaborative projects: UNIVERSELF (FP7), HOMESNET (Celtic) and ECOSCELLS (ANR).

Eitan Altman received the B.Sc. degree in electrical engineering (1984), the B.A. degree in physics (1984) and the Ph.D. degree in electrical engineering (1990), all from the Technion-Israel Institute, Haifa. In 1990 he further received his B.Mus. degree in music composition in Tel-Aviv university. Since 1990, Dr. Altman has been a researcher at INRIA (National research institute in computer science and control) in Sophia-Antipolis, France. He has been in the editorial boards of several scientific journals: Wireless Networks (WINET), Computer Networks (COMNET), Computer Communications (Comcom), J. Discrete Event Dynamic Systems (JDEDS), SIAM J. of Control and Optimization (SICON), Stochastic Models, and Journal of Economy Dynamic and Control (JEDC). He received the best paper award in the Networking 2006, in Globecom 2007 and in IFIP Wireless Days 2009 conferences, and is a coauthor of two papers that have received the best student paper awards (at QoFis 2000 and at Networking 2002). His areas of interest include networking, stochastic control and game theory. More information can be found at http://www-sop.inria.fr/members/Eitan.Altman/.