

Energy-Efficiency and Future Knowledge Tradeoff in Small Cells Prediction-Based Strategies

Matthieu De Mari

Email: matthieu.demari@supelec.fr

Emilio Calvanese Strinati

Email: emilio.calvanese-strinati@cea.fr

Merouane Debbah

Email: merouane.debbah@supelec.fr

Abstract—Predictive small cells networks and proactive resource allocation are considered as one of the key mechanisms for increasing the long-term energy-efficiency of communication networks. Learning techniques exploit repetitive patterns in human behavior to predict some future transmission contexts of the network. In this paper, we target to improve the energy efficiency of delay-tolerant transmissions by enabling flexibility in resource allocation with prediction-based strategies. We study the performance, in terms of energy efficiency of several scenarios of future knowledge ranging from zero to perfect knowledge of the future context, but also partial knowledge scenarios (short-term predictions, long-term statistics or partial knowledge). An iterative process, approaching the optimal strategies in each scenario, is described. In some cases, closed-form expressions of the optimal strategies to be implemented can be obtained and the performance in each scenario is computed. Our analytical and numerical results assess the potential benefit of exploiting the knowledge of the future in the case of a delay-tolerant transmission and show how the system may benefit from a provided piece of information about the future transmission context.

I. INTRODUCTION

Wireless networks are the communication bridge and bottleneck for future evolutions of the digital world. Operators struggle to support the massive data traffic growth in a sustainable and economical way [1]. In that sense, operators aim at reducing the networks power consumption, while maintaining a satisfying Quality of Service (QoS). To do so, small cells networks (SCNs) have been proposed as a valid candidate [2]. Several tradeoffs have also been identified [3], among which the latency-power efficiency tradeoff which lead to delay-tolerant transmission concepts. At the same time, recent advances on learning and data mining have shown that human behavior and future context for users in the network is highly and accurately predictable [4] [5]. As a consequence, recent works have then looked forward to coupling scheduling techniques with context predictions, leading to so-called proactive small cells networks.

Scenarios of proactive resource allocation and predictive wireless networks have been introduced and investigated in [6][7][8]. In these papers, the system is able to predict upcoming requests and can adapt its transmissions in order to limit its own outage probability. Significant diversity gains were analytically demonstrated and illustrate the significance of the potential benefit offered by knowledge of the future. In this paper and as in [9] [10], we investigate a scheduler for delay-tolerant transmissions, but consider several degrees

of knowledge of the future. For different assumptions on the knowledge of the future available, we show the performance of the system, when it implements several optimal and sub-optimal transmission strategies. However, in most cases, it appears that the closed-form expressions of the optimal scheduling policies and the performance of such systems can not be simply and clearly established. Instead, an iterative approach is considered to define, at each present time, the optimal power to be used.

In this paper, we propose several scenarios of downlink transmissions where a Base Station (BS) has to achieve data transmission of $Q(0)$ bits over block fading channels, with a deadline of T time slots. The system controls its instantaneous power of transmission p or equivalently can adapt its instantaneous rate to the present channel realisation h , the number of remaining time slots and the remaining data packet to be transmitted at the present time. The optimization problem consists in minimizing the cumulated power cost while satisfying a strong QoS requirement (a complete transmission of a packet of initial size $Q(0)$ before deadline T), for several different assumptions on the future context knowledge. We focus on defining the optimal strategies that the system implements, for different kinds of future context knowledge, ranging from a complete lack of knowledge to an ideal, complete and perfect knowledge of the future.

The investigated schedulers include scenarios where no knowledge of the future channel realisations is available to the system. We have also considered the case where a complete and exact knowledge is available to the system, which define the optimal performance bound. Schedulers optimal strategies and performances are also detailed in scenarios where partial and hybrid knowledge is available. This includes statistical long-term knowledge, but also partial, accurate and short-term estimations.

In this paper, our contribution lies in analysing the transmission policies and their performance in terms of energy efficiency, for several scenarios of future knowledge. This analysis leads to interesting insights, showing the potential significant benefit of transmission schedulers, coupling both delay-constrained settings and future context knowledge. It also shows how the system may benefit from elements of future knowledge, even statistical or partial.

In Section II, we define both the system model and the optimization problem we consider throughout the paper. In

Section III, an iterative process, allowing to approach the optimal transmission strategies for any given future knowledge is described. In Section IV, we define the different types of future knowledge that we consider in this paper and define, when possible the closed-form expression for both their power strategies and global power performances. Finally, we propose numerical simulations, compare and discuss each scheduler performance in Section V.

II. SYSTEM MODEL AND OPTIMIZATION PROBLEM

In this paper, we consider a simple downlink transmission model (similar to [9] [10]), consisting of one BS and one user. The BS is enforced to transmit a data packet to its associated user, within a limited number of time slots T . In the following, we denote the index of any time slot by $t \in 1, 2, \dots, T$.

In this context and as depicted in Figure 1, we consider for each time slot, uncorrelated block fading channels, in power units, $h(t), t \in \{1, 2, \dots, T\}$, with values in \mathcal{H} . We consider that a minimum link quality is always guaranteed, denoted $\epsilon > 0$, i.e. that $\mathcal{H} =]\epsilon, \infty[^T$. This assumption will later appear necessary, when computing strategies for which we do not have future knowledge. We assume, that the channel realisations are random i.i.d processes, distributed according to PDFs $D_{real}^1(h), \dots, D_{real}^T(h)$.

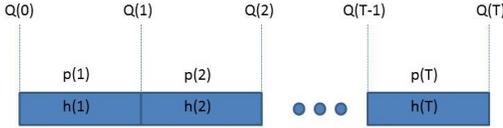


Fig. 1. Transmission scheme

We assume, that the channel can be perfectly estimated, at the transmitter, at the beginning of each time slot, and that it remains static for the complete duration of the time slot Δt . We also denote by $Q(t)$, the number of remaining bits at the end of time slot t . The initial amount of data, denoted $Q(0) > 0$, is known at the beginning of the first time slot. For simplicity, we assume that no other requests are allowed to enter the system until the end.

We assume that the BS can adapt its power of transmission $p = (p(t))_{t \in \{1, 2, \dots, T\}}$, where $p(t) > 0$ is the power level used at time slot t , which is defined at the beginning of each time slot t . The transmission rate matches the channel capacity, i.e. that the remaining packet size at the end of time slot t , $Q(t)$ decreases according:

$$\forall t \in \{1, T\}, Q(t) = Q(t-1) - B \cdot \log_2(1 + h(t)p(t)) \Delta t \quad (1)$$

Where Δt and B constants denoting the duration of the time slot and the bandwidth of the channel respectively.

We denote $\forall t, i \in \{1, \dots, T\}, i > t, D_i^t(h)$ as the prediction being made at the beginning of TS t on the channel realisation on TS i . In fact, $D_i^t(h)$ is an estimated PDF of the channel realisation $h(i)$, given to the system at TS t .

Within this context, we consider the following constrained optimization problem: the BS has to define a strategy that allows a complete transmission, of initial size $Q(0)$, before deadline T (achieving a complete transmission is then expressed by $Q(T) = 0$), at a minimal cumulated power cost. The problem can be formally rewritten as (2):

$$p^* = (p^*(1), p^*(2), \dots, p^*(T)) = \arg \min_p \sum_{k=1}^{k=T} p(k)$$

$$\text{s.t.}, Q(T) = Q(0) - \sum_{k=1}^{k=T} B \log_2(1 + h(k)p(k)) \Delta t = 0 \quad (2)$$

And $\forall t, i \in \{1, \dots, T\}, i > t, D_i^t(h)$ are known at the beginning of TS t .

In the following, we consider several assumptions on the degree of knowledge offered to the system, i.e. several and different predictions scenarios on PDFs $D_{i>t}^t(h)$. We first focus on how the system exploits such a given information about the future, to compute at the beginning of each TS t , the optimal power to be used, $p(t)$.

III. AN ITERATIVE APPROACH TO THE OPTIMIZATION PROBLEM

At the beginning of each TS t , the system is revealed the channel realisation for the TS t , denoted $h_r(t)$. Based on this present information and for a given knowledge of the future, represented by PDFs on the upcoming channel realisations ($D_{t+1}^t(h), \dots, D_T^t(h)$), the system has to define the optimal power $p^*(t)$ to be used for TS t , assuming the remaining packet to transmit until deadline T is of $Q(t-1)$.

The optimal power $p^*(t | h_r(t), Q(t-1))$ to be used at time t is defined at the beginning of TS t , as :

$$p^*(t | h_r(t), Q(t-1)) = \arg \min_p \left(p + \mathbb{E} \left[\sum_{k=t+1}^{k=T} p(k) | Q(t) \right] \right) \quad (3)$$

With $Q(t) = Q(t-1) - B \log_2(1 + h_r(t)p)$

And $\forall i \in \{t+1, \dots, T\}, D_i^t(h)$ are known.

The difficulty of the optimization process relies in being able to estimate the cost-to-go function $S(t+1 | Q(t)) = \mathbb{E} \left[\sum_{k=t+1}^{k=T} p(k) | Q(t) \right]$ at time t , since it depends on all the random channel realisations of the $T-t$ remaining time slots, ($h(t+1), \dots, h(T)$). However, the optimization (3) can be formulated sequentially with dynamic programming. In fact, it is a classic backward iteration, as in [11].

The optimal power to be used on TS T , $p^*(T | h, Q)$ if the channel realisation $h_r(T) = h$ and the remaining packet $Q(T-1) = Q$ can be easily computed, as the power necessary to complete the transmission on the last time slot T , i.e.

$$p^*(T | h, Q) = (2^{(\frac{Q}{B\Delta t})} - 1) \frac{1}{h} \quad (4)$$

At $t = T-1$, the system can compute $S(T | Q) = \mathbb{E} [p(T | Q)]$, as:

$$S(T | Q) = \int_{h \in \mathcal{H}} D_T^t(h) p^*(T | h, Q) dh \quad (5)$$

Which leads to:

$$S(T | Q) = (2^{\frac{Q}{B\Delta t}} - 1) \mathbb{E} \left[\frac{1}{h(T)} \right] \quad (6)$$

Note that $\mathbb{E} \left[\frac{1}{h(T)} \right]$ is not finite if the channels have Rayleigh fading (i.e. $h(T)$ is exponentially distributed), unless the channel set for possible channel realisations \mathcal{H} is truncated as described in [13]. In fact, this is one of the reasons why we have considered a minimal channel realisation $\epsilon > 0$ and considered the set of admissible channel realisations \mathcal{H} to be $]\epsilon, \infty[^T$.

Based on the previous results, we can solve the optimization problem 3, via the following iterative process. For k from $T-1$ to t , we sequentially compute:

$$p^*(k | h, Q) = \arg \min_p (p + S(k+1 | Q')) \quad (7)$$

With $Q' = Q - B \log_2(1 + p * h)$. This is a one-dimensional convex optimization problem, in p , which is easy to solve. Based on this, we can compute $S(k | Q)$, as:

$$S(k | Q) = \int_{h \in \mathcal{H}} D_k^t(h) p^*(k | h, Q) dh + \int_{h \in \mathcal{H}} D_k^t(h) S(k+1 | Q') dh \quad (8)$$

Where $Q' = Q - B \log_2(1 + p^*(k | h, Q)h)$.

The iterative process is repeated until $p^*(t | h, Q)$ is defined. For a given present realisation of the channel $h_r(t)$ and assuming the remaining packet size is $Q(t-1)$, we can solve the optimization 3, by defining the optimal power to be used on time slot t , $p^*(t)$, as:

$$p^*(t) = p^*(t | h_r(t), Q(t-1)) \quad (9)$$

At the beginning of each time slot t , we can compute the optimal power $p^*(t)$, based on the remaining packet size $Q(t-1)$, the present and revealed channel realisation $h_r(t)$ and take into account the future channel estimations $(D_i^t(h))_{i>t}$ available at time t .

We now focus on the different scenarios of future knowledge and look forward to define, when it is possible, the closed-form expressions of the optimal transmission policies and their global performance.

IV. SCENARIOS OF FUTURE KNOWLEDGE: OPTIMAL SCHEDULING AND PERFORMANCE

A. Omniscient Scheduling

We define the optimal power strategy p_{om} , as the optimal strategy solving (2) when the system is given perfect knowledge of the future channel realisations h_r at $t = 0$. In this configuration, we assume the system has a perfect knowledge of the future channel realisations, i.e., the predictions PDFs $D_i^t(h)$ are defined as:

$$\forall i, t \in \{1, \dots, T\}, i > t, D_i^t(h) = \delta_{h=h_r(i)} \quad (10)$$

Where $\delta_{h=h_r(i)}$ is a Dirac distribution centered on $h_r(i)$.

It turns out that the iterative process we defined in Section III, for which we have defined $D_i^t(h)$ as in Eq. 10, leads to the exact same power strategy p_{om} , defined by Time-Water-Filling, in Proposition IV.1.

Proposition IV.1. *The power strategy p_{om} consists of a time water-filling:*

$$\forall t \in 1, \dots, T, p_{om}(t) = \max(0, \mu - \frac{1}{h(t)}) \quad (11)$$

Where μ is the unique water-level, verifying:

$$Q(0) = \sum_{t=1}^T B \log(1 + h(t)p_{om}(t)) \Delta t \quad (12)$$

Proof. Elements of proof can be found in [9] and [14] \square

The strategy p_{om} gives the optimal performance bound, that can be achieved for any given future realisation of the channel $h = (h(1), \dots, h(T))$ and a constrained data transmission, with initial size $Q(0)$ and deadline T . The closed-form expression of the power strategies is not easy to define, because it depends on the number of time slots active for transmission $\mathcal{N}(Q(0), h)$, defined as:

$$\mathcal{N}(Q(0), h) = \text{card} \{p_{om}(t) > 0 \mid t \in \{1, \dots, T\}, Q(0), h\} \quad (13)$$

Assuming, the channel realisations happen according to PDFs $D_{real}^1(h), \dots, D_{real}^T(h)$, computing $\mathbb{E} \left[\sum_{k=1}^{k=T} p(k) \mid Q(0) \right]$ in such a scenario is also complicated, since it depends on $\mathcal{N}(Q(0), h)$ as well. An asymptotic approach has however been considered to approximate the system performance, in cases where $\frac{Q(0)}{B\delta t}$ is large [14].

B. Zero-knowledge optimal strategy I: Worst-Case Scenario

We define by *zero knowledge* strategy, the power strategy p_{zk} that would be implemented, if the system is given the least possible information about its future context. In this scenario, the next future channel realisations remain unknown to the BS, until the beginning of each time slot, where each channel is revealed. A first possible approach would consist to define the power $p_{zk}(t)$ to be used at the beginning of each time slot t , for a revealed channel realisation $h(t)$ on time slot t , by assuming that the future realisations will lead to the worst possible configuration. This power allocation scheme is then completely reactive, in the sense that it is not able to exploit any information about the future, but has to deal with the channel information revealed at the present time. Such a strategy leads to the best worst-case performance.

In this context, the best power strategy $p_{zk}(t)$ to be implemented on time slot t corresponds to the first element of the optimal power strategy $p = (p(t), \dots, p(T))$, solving the following min-max optimization problem.

$$\min_p \max_{(h(t+1), \dots, h(T)) \in \mathbb{H}^{(T-t)}} \left(\sum_{k=t}^{k=T} p(k) \right)$$

$$\text{s.t.}, Q(t-1) = \sum_{k=t}^T B \cdot \log(1 + h(k)p(k)) \Delta t \quad (14)$$

The worst possible future configuration can be easily defined as the case where all the unknown future channel realisations ($h(t+1), \dots, h(T)$) take the smallest value ϵ . In this case, we have $\forall i, t \in \{1, \dots, T\}, i > t, D_i^t(h) = \delta_{h=\epsilon}$ and the optimization problem (14) can be rewritten as:

$$\min_p \left(\sum_{k=t}^{k=T} p(k) \right)$$

$$\text{s.t.}, Q(t-1) = \sum_{k=t}^T B \cdot \log(1 + h(k)p(k)) \Delta t$$

$$h(t) = h_r(t) \in \mathcal{H} \text{ known, and } \forall k \in \{t+1, \dots, T\}, h(k) = \epsilon \quad (15)$$

The optimization problem leads to the following time water-filling solution:

$$\forall k \in \{t, \dots, T\}, p(k) = \begin{cases} \max(0, \mu - \frac{1}{h(t)}) & \text{if } k = t \\ \max(0, \mu - \frac{1}{\epsilon}) & \text{else.} \end{cases} \quad (16)$$

Where μ is the unique water-level satisfying:

$$Q(t-1) = \sum_{k=t}^T B \cdot \log(1 + h(k)p(k)) \Delta t$$

Proposition IV.2. *The power strategy p_{zk} is then defined as:*

$$p_{zk}(t) = \left(2^{\frac{Q(t-1) - Q(t)}{B\Delta t}} - 1 \right) \frac{1}{h(t)} \quad (17)$$

Where the packet remaining at time t , $Q(t)$ decreases with t and is given by:

$$\begin{aligned} Q(t) &= \left(\frac{T-t}{T-t+1} \left(Q(t-1) - B \log_2 \left(\frac{h(t)}{\epsilon} \right) \right) \Delta t \right)^+ \\ &= \left(\frac{T-t}{T} Q(0) - \sum_{i=1}^t \frac{T-t}{T-i+1} B \log_2 \left(\frac{h(i)}{\epsilon} \right) \Delta t \right)^+ \end{aligned} \quad (18)$$

Proof. For elements of proof, refer to [14] \square

Within this framework, we show, that a BS can smartly exploit the limited knowledge, consisting of each channel realisation revealed at the beginning of each time slot, and adapt its power strategy in a reactive way. However, the system pessimistically assumes the worst scenario for the future realisations of the channel. Because of this, the strategy p_{zk} appears unable to fully exploit the latency/power efficiency tradeoff and, as a consequence, the global performance, in terms of energy-efficiency, of the *zero knowledge* strategy p_{zk} is poor compared to the omniscient one p_{om} . Computing analytically the expected performance of this strategy leads to serious complications, which is the reason why, we will not develop it extensively in this paper.

C. Zero-knowledge optimal strategy II: Equal-bit Scenario

Another power strategy one could implement, in a scenario where no knowledge of the future is available, is the *equal-bit* strategy. Basically, the scheduler transmits $\frac{Q(0)}{T}$ bits during each time slot, no matter what the present or future channel realisations might be.

Proposition IV.3. *The power strategy p_{eb} is then defined as:*

$$p_{eb}(t) = \left(2^{\frac{Q(0)}{TB\Delta t}} - 1 \right) \frac{1}{h_r(t)} \quad (19)$$

This scheduler has an expected total power consumption $\mathbb{E} \left[\sum_{k=1}^{k=T} p(k) \mid Q(0) \right]$, which can be immediately defined as:

$$\mathbb{E} \left[\sum_{k=1}^{k=T} p(k) \mid Q(0) \right] = \left(2^{\frac{Q(0)}{TB\Delta t}} - 1 \right) \sum_{i=1}^T \int_{h \in \mathcal{H}} D_{real}^i(h) dh \quad (20)$$

This strategy is able to take into account the number of extra time slots available for transmission, but does not consider any future channel predictions. As a consequence, the system is unable to smartly distribute the power over the time slots with good channel realisations, while avoiding transmissions on the time slots with poor channel realisations. In the end, its global energy efficiency performance is poor compared to a scenario where such a future knowledge, even partial, is provided to the system.

D. Statistical Knowledge Scenario

A first kind of knowledge, that the system can access, is a *statistical knowledge* about the future channel realisations. In this scenario, we assume that the system is not able to tell exactly, what the future channel realisations will be. Instead, it can access the exact long-term channel statistics ($D_{real}^i(h)_{i \in \{1, \dots, T\}}$), which means that:

$$\forall i, t \in \{1, \dots, T\}, i > t, D_i^t(h) = D_{real}^i(h) \quad (21)$$

The iterative process we defined in Section III, uses the available predictions on channel realisations $D_i^t(h)$, $\forall i, t \in \{1, \dots, T\}, i > t$. This strategy allows to take into account the possible channel realisations for the upcoming time slots. This way, the present power strategy p_{st} , whose elements are computed at the beginning of each time slot, is able to take into account both the number of remaining time slots $T-t$ and some piece of statistical information about what the remaining future channel realisations might be.

Note that, we have considered, that the system was not able to adjust and refine its statistical estimations, at the beginning of each time slot. In fact, we could model such a scenario, where our prediction $D_i^t(h)$ gets more and more accurate, as long as the present time slot t gets closer to the time slot i . In that case, when we get closer to the future time slot i , $D_i^t(h)$ converges to the best predictor one could make about the channel realisations on time slot i , $\delta_{h=h_r(i)}$. Such a scenario of future knowledge is not investigated throughout this paper, but will be in an upcoming journal paper.

E. Two Scenarios of Short-Sighted Knowledge

An other scenario of future knowledge that the system can access consists of a *short-sighted knowledge*. In this scenario, we assume that the system is able to estimate accurately the present channel realisation, at the beginning of each time slot, as well as a few upcoming channel realisations. In fact, there are a lot of channel models, that are able to predict the channel realisations for upcoming time slots, by taking into account the present channel realisations [12].

In this section, we assume that at the beginning of each time slot, the system is able to predict exactly the channel realisations for the upcoming $K \in 1, \dots, T$ time slots. This means, that, at the beginning of time slot t , the prediction provided to the system is:

$$\forall i \in \{t, \dots, \min(T, t + K - 1)\}, i \geq t, D_i^t(h) = \delta_{h=h_r(i)} \quad (22)$$

The remaining predictions, can be modelled according to two scenarios:

- *K-Short-Sighted with Statistical Knowledge on the remaining time slots*: we assume that the remaining predictors are defined as in the statistical Section IV-D, i.e.

$$\forall i \in \{1, \dots, T\}, i \geq t + K, D_i^t(h) = D_{real}^i(h) \quad (23)$$

- *K-Short-Sighted with Zero Knowledge on the remaining time slots*: we assume that the remaining predictors are defined as in the Zero Knowledge I Section IV-B, i.e.

$$\forall i \in \{1, \dots, T\}, i \geq t + K, D_i^t(h) = \delta_{h=\epsilon} \quad (24)$$

In both cases, our iterative method allows to compute the optimal power $p^*(t)$ to be used at the beginning of each time slot t .

V. NUMERICAL PERFORMANCE ESTIMATION AND COMPARISONS

A. Simulation Parameters

In order to numerically evaluate the expectation of the consumed power, we ran Monte-Carlo simulations over $M_c = 1000$ iterations for arbitrary varying values of $\frac{Q(0)}{B\Delta t}$ and T . The performances of the schedulers are numerically estimated and compared, in a scenario of block-fading Rayleigh channels with parameter $\lambda = 1$. The simulation parameters are listed in Table I, below.

TABLE I
SIMULATION PARAMETERS

Parameter	Parameter Value
$\frac{Q(0)}{B\Delta t}$	ranging from 1 to 200
Number of time slots T	25, as in [10]
Number of MC iterations M_c	1000
Channels	Block fading with i.i.d. trunc. Rayleigh(1)
Channel set \mathcal{H}	$]\epsilon, \infty[$
ϵ	0.1

B. About the Significance of a Performance Gain

Our first concern when considering schedulers that are able to take into account predictions about the future transmission context is to determine how significant the potential energy performance gains might be, between:

- the performance of the most efficient power strategy we can implement when the system is not given any information about its future transmission context,
- the performance of a scheduler taking into account a statistical knowledge of the future
- and the best performance the system can access, if it is provided a complete exact knowledge of the future.

To do so, we first considered simulations with $T = 25$ time slots and $\frac{Q(0)}{B\Delta t} = 100$. Let us now consider

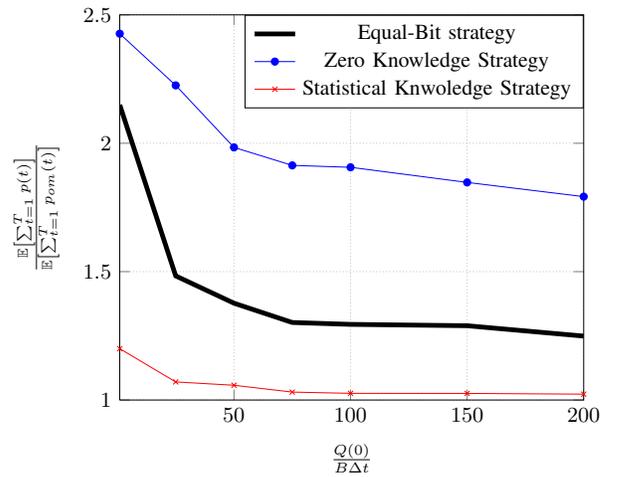


Fig. 2. Energy Performance vs. $\frac{Q(0)}{B\Delta t}$ - $\frac{Q(0)}{B\Delta t}$ ranging from 1 to 100 and $T = 25$.

As expected from the analytical investigation, it appears that the *zero knowledge worst-case* scenario is unable to benefit from the T time slots, since it is assuming the worst possible realisation of the future at the beginning of each time slot. It turns out that the scheduler rushes the transmission of the packets, completing the transmission notably before the deadline. Its global power efficiency is then poor compared to the two other strategies. We also know, from our analytical study, that the *equal-bit* scheduler is not able to identify the good and bad channel realisations. As a consequence, it is unable to smartly distribute its transmission powers among the most appropriate channels. In the end, it results in a poor power efficiency performance compared to the *omniscient* scenario. In the following, we considered the ratio between the performance of the transmission strategy and the total power of the omniscient strategy, $\frac{\mathbb{E}[\sum_{t=1}^T p(t)]}{\mathbb{E}[\sum_{t=1}^T P_{om}(t)]}$, as our energy efficiency performance criterion. The average energy performances of each scheduler are depicted on Figure 2, for a values set $\frac{Q(0)}{B\Delta t}$, when $T = 25$. It turns out that the *statistical knowledge* scheduler outperforms both the Equal-Bit scheduler and the *zero knowledge worst-case* scheduler.

Also, when no information about the future is available at all, the *equal-bit* scheduler turns out to be a more energy efficient scheduler, compared to the *zero knowledge worst-case* scheduler. We observe that there is a significant potential gain between the best performing scheduler when no information about the future is available and the performances of both the *statistical knowledge* scheduler and *omniscient* scheduler. This confirms, that acquiring a future knowledge, even partial or statistical, appears worth the effort [15], because we observe a significant potential energy efficiency improvement.

C. Performance analysis

In this section, we investigate the performance of the two short-sighted schedulers we proposed in Section IV-E. Moreover, in this section, we present a comparison of all five strategies performances.

First, let us define the $\alpha_z(p^*)$ and $\alpha_s(p^*)$ criteria defined as:

$$\alpha_z(p^*) = \frac{\mathbb{E} \left[\sum_{t=1}^T p_{zk}(t) \right] - \mathbb{E} \left[\sum_{t=1}^T p^*(t) \right]}{\mathbb{E} \left[\sum_{t=1}^T p_{zk}(t) \right] - \mathbb{E} \left[\sum_{t=1}^T p_{om}(t) \right]} \quad (25)$$

$$\alpha_s(p^*) = \frac{\mathbb{E} \left[\sum_{t=1}^T p_{st}(t) \right] - \mathbb{E} \left[\sum_{t=1}^T p^*(t) \right]}{\mathbb{E} \left[\sum_{t=1}^T p_{st}(t) \right] - \mathbb{E} \left[\sum_{t=1}^T p_{om}(t) \right]} \quad (26)$$

On Figure 3, we show the average α_z criterion value for optimal strategies in *K short-sighted scenarios with zero knowledge*, for values of K ranging from 1 to T . We have also represented the performance of Equal-Bit scheduler. As expected, the system has the same performance of the *zero knowledge* scheduler when $K = 1$ and has the same performance of the *omniscient* scheduler when $K = T$. From the previous section, we know that when the system is not given any information about the future, the *equal-bit* scheduler is the best performing strategy. It appears that the *K short-sighted strategy with zero knowledge* outperforms the Equal-Bit scheduler for values of $\frac{K}{T} \geq 0.15$.

The same way, we have considered the scenarios where *statistical knowledge* is available as a complement to the K upcoming estimated channel realisations, namely the *K-Short-Sighted with Statistical Knowledge on the remaining time slots*. On Figure 4, we show the average α_s criterion value for optimal strategies in *K short-sighted scenarios with statistical knowledge*, for several values of K ranging from 1 to T . We have also represented the performance of the *statistical knowledge*, *equal-bit* and *omniscient* schedulers. As expected, the system has the same performance of the *statistical knowledge* scheduler when $K = 1$ and has the same performance of the *omniscient* scheduler when $K = T$.

On both figures, it appears that the performance of the system increases when $\frac{K}{T}$ increases as well. Also, the larger K becomes, the less significant the extra performance gain is. As a conclusion, it appears that providing a *K short-sighted knowledge* to a system with no other information

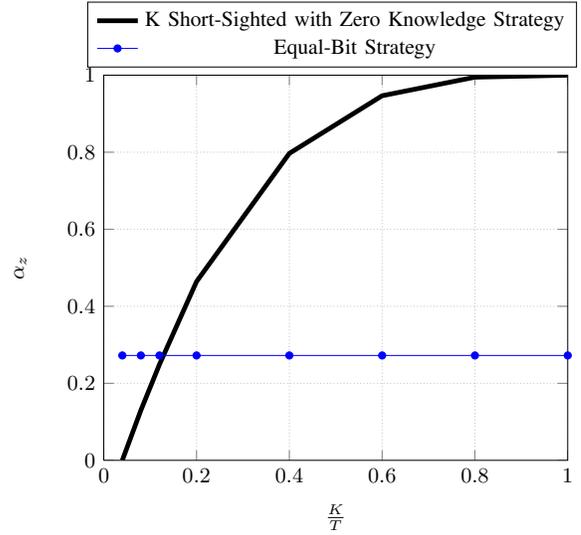


Fig. 3. α_z criterion vs. $\frac{K}{T} - \frac{Q(0)}{B\Delta t} = 100$ and $T = 25$.

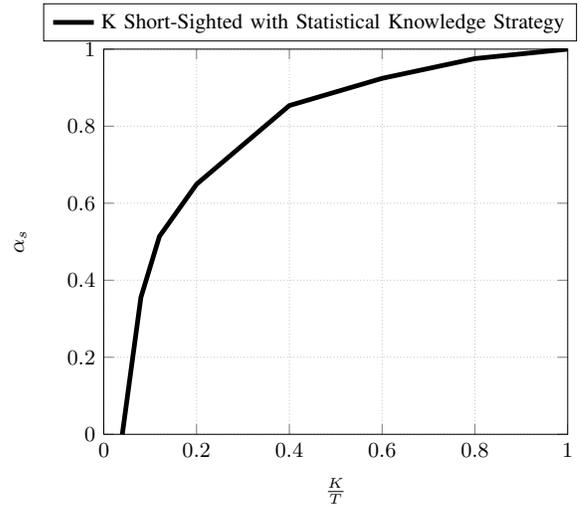


Fig. 4. α_s criterion vs. $\frac{K}{T} - \frac{Q(0)}{B\Delta t} = 100$ and $T = 25$.

about the future is highly benefic to the system, even when K is small. In fact, it suffices that $K \geq 0.15T$ so that the information becomes benefic to the system, leading to an average performance greater than the average performance of the *equal-bit* scheduler.

When Statistical knowledge is also available to the system, a short-term prediction capability improves even more the performance of the designed scheduler. Its average performance converges, when $K \rightarrow T$, with a similar behavior to the one described before, to the optimal bound, i.e. the average performance of the omniscient scheduler.

Finally, we discuss the performance of several future knowledge scenarios and schedulers. The following schedulers have been considered:

- Zero Knowledge Scheduler
- Equal-Bit Scheduler

- Statistical Knowledge Scheduler
- K Short-Sighted, with Zero Knowledge Scheduler ($K = 2, 3, 5$)
- K Short-Sighted, with Statistical Knowledge Scheduler ($K = 2, 3$)

Figure 5 shows the ratio between the average energy performance of each strategy and the average performance of the omniscient scheduler for several values of $\frac{Q(0)}{B\Delta t}$ and $T = 25$.

It appears, as expected, that the performance of the K Short-Sighted, with Zero Knowledge scheduler is better than the one of the equal-bit scheduler, when $\frac{K}{T} \geq 0.15$, and we observe that the performance of the K Short-Sighted, with Zero Knowledge scheduler improves as long as K increases, with decreasing gain every time. The performance of the Statistical Knowledge scheduler is quite close to the optimal one, as demonstrated also in Section V-B. Adding a K Short-Sighted Knowledge to the Statistical Knowledge contributes to improve even more the performance of the scheduler, which rapidly tends to the optimal performance achieved by the Omniscient scheduler.

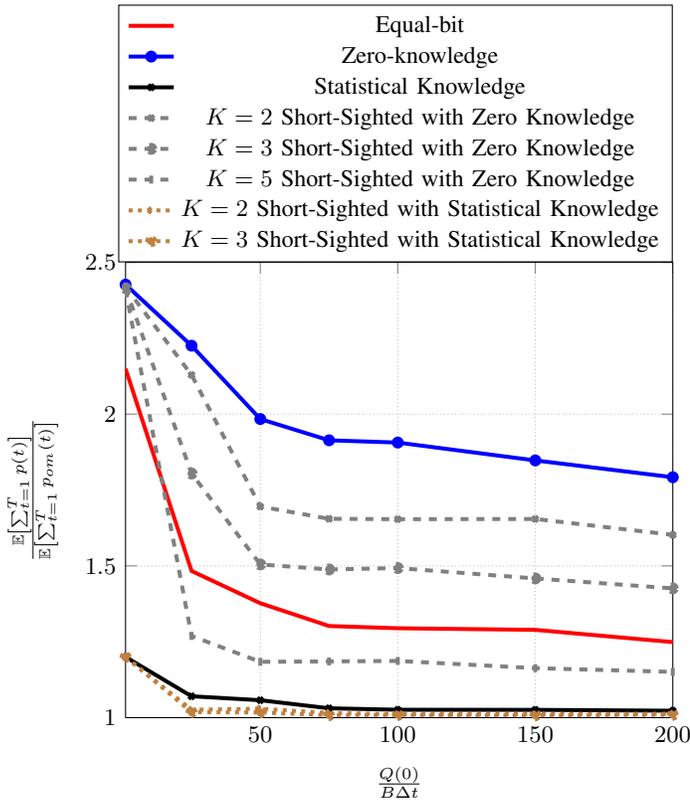


Fig. 5. Energy Performance vs. $\frac{Q(0)}{B\Delta t}$ - $\frac{Q(0)}{B\Delta t}$ ranging from 1 to 200 and $T = 25$.

VI. CONCLUSION AND DISCUSSION

In this paper, we investigate the tradeoff between the latency T , the future context knowledge and the energy performance of the system for delay-tolerant transmissions. For several identified scenarios of future knowledge ranging from zero

to complete knowledge, we propose an iterative algorithm to approach the optimal instantaneous power strategies. It appears that the system may greatly benefit of any piece of future knowledge information, such as channel statistics or short-term estimations. The system can then implement scheduler, for which power performance appears greatly improved, compared to scenarios where no knowledge is available.

As future work, such results will be discussed. For example, we have not assumed any cost related to the learning capability: one could discuss the actual cost of learning and confront it to the actual energy-efficiency improvement, that the system can access if it is granted a certain knowledge of its upcoming future transmission context. New hybrid strategies able to take into account both short-sighted predictions and partial/statistical knowledge are also being studied.

ACKNOWLEDGMENT

This research has been supported by the ERC Starting Grant 305123 MORE (Advanced Mathematical Tools for Complex Network Engineering)

REFERENCES

- [1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 20122017", *Cisco White Paper*
- [2] Hoydis, J., Kobayashi, M., & Debbah, M. (2011). "Green small-cell networks". *Vehicular Technology Magazine, IEEE*, 6(1), 37-43.
- [3] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental Tradeoffs on Green Wireless Networks", *Communications Magazine, IEEE*, Vol. 49, 2011
- [4] M. De Domenico, A. Lima and M. Musolesi, "Interdependence and Predictability of Human Mobility and Social Interactions", *Mobile Data Challenge*, 2012
- [5] C. Song, et al., "Limits of Predictability in Human Mobility", *Science* 327, 2010
- [6] H. El Gamal, J. Tadrous and A. Eryilmaz, "Proactive Resource Allocation: Turning Predictable Behavior into Spectral Gain", *Allerton 2010*
- [7] J. Tadrous, A. Eryilmaz and H. El Gamal, "Proactive Resource Allocation: Harnessing the Diversity and Multicast Gains", *submitted to IEEE Transactions on Information Theory*
- [8] J. Tadrous, A. Eryilmaz and H. El Gamal, "Proactive Resource Allocation in Cognitive Networks", *ASILOMAR 2011*
- [9] J. Lee and N. Jindal, Energy-efficient scheduling of delay constrained traffic over fading channels, *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 18661875, Apr. 2009.
- [10] J. Lee, and N. Jindal, Asymptotically optimal policies for hard-deadline scheduling over fading channels, *Submitted to IEEE Transactions on Information Theory*, June 2009.
- [11] Boyd, S. P., & Vandenberghe, L. (2004). "Convex optimization". *Cambridge university press*.
- [12] Hellings, C., Kiefer, P., & Utschick, W., "Rate Prediction and Receding Horizon Power Minimization in Block-Fading Broadcast Channels"
- [13] D. Tse and P. Viswanath, "Fundamentals of Wireless Communication", *Wiley series in telecommunications*, 2005
- [14] M. De Mari, E. Calvanese Strinati and M. Debbah, "Elements of Proof - Energy-Efficiency and Future Knowledge Tradeoff in Prediction-Based Strategies". Available at: <https://docs.google.com/file/d/0B-8455WISEiOcTN3U1IyRjc4am8>
- [15] Proebster, M., Kaschub, M., & Valentin, S. (2011, June). "Context-aware resource allocation to improve the quality of service of heterogeneous traffic". *In Communications (ICC), 2011 IEEE International Conference on* (pp. 1-6). *IEEE*.