Optimal Sensing Policies for Smartphones in Hybrid Networks: A POMDP Approach

Mohammed Raiss-el-fenni[†] , Rachid El-Azouzi[†] , Daniel S Menasche[◇], Yuedong Xu[†]

***Abstract*—The ubiquity of mobile devices is fostering a paradigm shift in the realm of Internet services. Consider, for instance, mobile users of social networks, that require frequent updates through small messages from their friends. If a user activates his mobile device and has a contact opportunity with an access point, an update can be received at the expense of monetary and energy costs. Thus, users face a tradeoff between such costs and the utilities of the messages received. The goal of this paper is to show how a user can cope with such a tradeoff, by deriving optimal *sensing policies*. A sensing policy consists of deciding, based on the age of the last message received and the *belief* about the future availability of a WiFi access point, whether to activate the mobile device or not. Alternatively, users may also decide to use 3G technology to receive updates, which provides broader coverage at the expense of higher monetary costs and lower bandwidth. To address the tradeoff faced by the users, we propose an analytical model based on a Partially Observed Markov Decision Process (POMDP) with an average reward criterion. Using the proposed model, we show properties of the optimal sensing policy. The applicability of the model and of the derived policy is illustrated through numerical case studies.**

I. Introduction

The ubiquity of mobile devices is fostering a paradigm shift in the realm of Internet services. Facebook alone counts with up to 800 million users, 41% of whom frequently access the social network through smartphones, and more than 475 mobile operators globally work to deploy and promote Facebook mobile products. Despite the fact that users are free to create accounts and upload content at sites like Facebook, they are faced with tradeoffs while deciding when and how to access the network from mobile devices.

As the population of users connected to social networks and sites alike keeps growing, status updates and notifications are generated at increasingly higher rates. As the time between updates decreases, users are faced with novel challenges. Mobile users accessing the network from mobile devices consume energy from limited batteries and their access might be subject to fees. We consider users that can access Internet through a hybrid network, which supports WiFi and 3G. While the 3G network provides broader coverage [4], its usage requires a subscription to an operator and its monetary and energy costs are significantly higher than WiFi.

[†]CERI/LIA, University of Avignon, 74 rue Louis Pasteur, 84029 AVIGNON Cedex 1, email: mohammed.raiss@etd.univ-avignon.fr

[†]CERI/LIA, University of Avignon, 74 rue Louis Pasteur, 84029 AVIGNON Cedex 1, email: rachid.elazouzi@univ-avignon.fr

[◇]University of Massachusetts, 140 Governors Drive, Amherst, email: adoc@cs.umass.edu

Let the *age* of a content[1] be the duration of time since the content was updated ($i.e.$, received) for the last time by a user. If a user is in the range of a service provider (WiFi access point or 3G antenna) and activates his mobile device, an update is received and the age of the message is reset to one, at the expense of a monetary and energy costs. Thus, users face a tradeoff between the costs and their *message utilities*. To cope with such a tradeoff, users decide, based on the availability of an WiFi access point (AP) and the age of the message, whether to activate the mobile device or not, and which technology to use (WiFi or 3G).

The main objective of this paper is to derive the optimal activation and access policy for a mobile user. We refer to a policy which determines activation decisions as a function of message ages and the availability of access points as a *sensing policy.*

Given the scenario outlined above, we pose the following question: what is the users optimal sensing policy? In answering this question, we make the following contributions.

Model formulation: we propose an analytical model to capture the tradeoff faced by mobile users while receiving their content updates. Our model is based on the framework of partially observed Markov decision processes (POMDP), wherein the state of a user comprises the age of its message as well as the belief about the chance of meeting an access point in the upcoming time slot.

Optimal policy: using the proposed model, we show several properties satisfied by optimal policies. In particular, we show that there exists a threshold $\lambda^\star$ such that if the user's belief about the opportunity of contacting an AP is smaller than $\lambda^\star$, it is optimal to remain idle. We establish monotonicity properties of the POMDP value function [14], and use them to get a closed-form expression for the threshold $\lambda^\star$ as a function of the system parameters. We also determine conditions on the age of the messages under which it is beneficial to access the network through WiFi as opposed to relying on the more costly 3G. The optimal policy can be determined using value iteration, and is simple enough to be easily implemented in off-the-shelf smartphones.

The organization of the remainder of this paper is as follows. In the next section we describe the mobile user model II. Section III formalizes the problem statement and Section IV presents the partially observed Markov decision process framework. In Section V we use the proposed model to establish the structural properties of an optimal policy.

[1]In this paper, the terms content and message are used interchangeably.

VALUETOOLS 2012, October 09-12, Cargèse, France

DOI 10.4108/valuetools.2012.250332

Section VI presents some numerical illustrations, related work is presented in Section VII and Section VIII summarizes our observations and concludes the paper.

II. Why Sensing Policies for Aging Control?

The goal of a sensing policy is to control the aging of the messages so as to reduce energy and monetary costs. Whereas the energy reduction is clearly of interest to users, monetary cost reduction impacts both users and providers.

The increasing demand for mobile Internet access is creating pressure on the service providers, whose limited spectrum might not be sufficient to cope with the demand. To deal with such pressure, some wireless providers are offering incentives to subscribers to reduce their 3G usage by switching to WiFi. This is beneficial not only to reduce the pressure over the 3G spectrum, but also for monetary reasons, given that WiFi technology is less expensive than 3G.

Due to the aforementioned reasons, open WiFi access points are becoming increasingly popular. Open WiFi access points motivate the sensing policies described in this paper, as users encounter such access points in an ad hoc fashion. Alternatively, users can also use proprietary WiFi access point infrastructure installed by service providers. In both cases, random factors such as fading and user speed determine the availability of the WiFi access points, which will vary in time and space.

A. Partially Observed Markov Decision Processes

We propose a general model that allows us to study the impact of energy costs, prices, utility of messages and their age on the sensing policy. Our problem is formulated as a partially observable Markov decision process (POMDP) [10],[13],[6]. *The POMDP accounts for the fact that contacts between WiFi access points and users occur in an unpredictable way (for the reasons pointed out above), but not uniformly at random (as illustrated in Section II-B).* Our model allows us to naturally consider the effect of the actions at a given time slot on the future states of the system, and allows us to derive the structure of the optimal sensing policy.

B. Access Point Contacts Are Unpredictable But Correlated

In this section we use traces collected from the UMass Amherst DieselNet [5] to study the distribution of contact opportunities between mobile users and access points. Mobile users considered here are passengers and drivers of buses. To characterize the update opportunities experienced by users, we analyze contacts between buses and access points at the UMass campus. Each bus scans for connection opportunities with APs on the road, and when found, connects to the AP and records the duration of the connection [5]. We assume that when the mobile devices are active, scans for access points occur every σ seconds. The scan is terminated once an access point is found. It was empirically determined that a scanning frequency of $1/20$ seconds yields a good balance between efficiency and low energy expenditure [15], so we considered $\sigma = 15$ and $\sigma = 20$ in our study. An access point is considered useful once it is scanned in two consecutive intervals of σ seconds. Henceforth, we refer to contact opportunities that last at least σ seconds simply as *contacts*.

Figure 1(a) (resp., Figure 1(b)) shows the CDF of the probability of a contact followed by no contact (resp., no contact followed by no contact). Each bus shift is divided into time slots of five minutes. If there is a contact in a given slot, we mark the slot as a useful slot. For each bus shift we generate a string of zeros and ones, corresponding to useful and non-useful slots, respectively. Then, we compute, for each bus shift, the fraction of ones followed by zeros (Figure 1(a)) and the fraction of zeros followed by zeros (Figure 1(b)). A point (x, y) in Figure 1(a) (resp., Figure 1(b)) represents the fact that at a fraction y of the bus shifts the probability of no contact at slot $t+1$ given a contact (resp., no contact) at slot t was smaller than x.

Figure 1(a) indicates that the median of the probability of a contact being followed by no contact is roughly 0.5, for $\sigma = 15$ and $\sigma = 20$. The probability is well approximated by a uniform distribution, in the range of [0.2,0.6]. Figure 1(b), in contrast, shows that the median of the probability of no contact being followed by no contact is roughly 0.45 and 0.55 for $\sigma = 15$ and $\sigma = 20$, respectively. This indicates that even though access point contacts are unpredictable, there is correlation among them, which we capture in the formulation presented in Section III and the corresponding POMDP model introduced in Section IV.

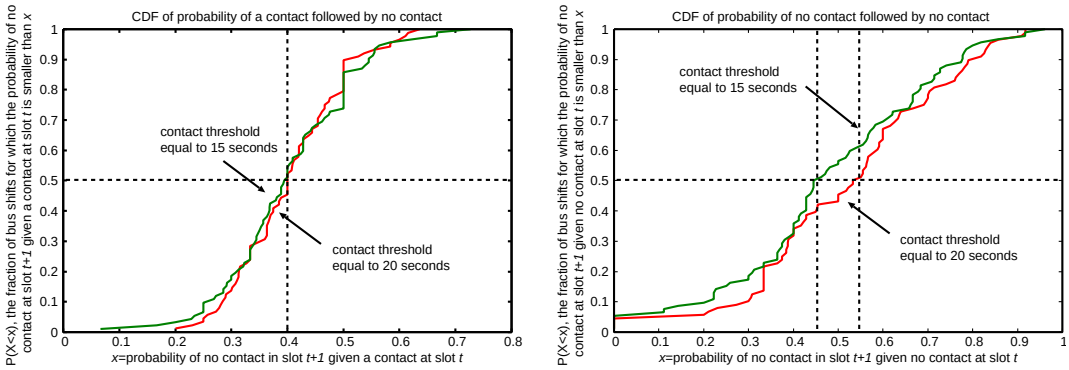
III. Model

We consider a time slotted hybrid wireless network where mobile users receive update messages from a service provider using either WiFi access points or 3G antennas. The availability of an access point is determined by some random factors like attenuation (fading) and user speed. The availability of an access point is modeled by a time homogeneous discrete Markov process $\{s(t) : t \geq 0\}$, $s(t) \in \{0, 1\}$, where $s(t) = 1$ means that an access point is available and $s(t) = 0$ means that no access point is available at time t. The transition probabilities of the access point are denoted by $P(s'|s) = P(s(t) = s'|s(t-1) = s)$. Let β (resp., α) be the probability that no contact (resp., a contact) is followed by a contact. Let P be the access point availability transition matrix,

$$P = \begin{pmatrix} 1-\beta & \beta \\ 1-\alpha & \alpha \end{pmatrix}$$

The transition probabilities can be determined based on the statistics of the service provider, as illustrated in Section II-B. In this paper we assume that the transition matrix P is known by mobile users. In practical scenarios, learning methods such as rate estimators and transition matrix estimators [18] allow mobile users to estimate the transition probabilities. Let $\pi(0)$ and $\pi(1)$ be the steady state probability of no contact and the steady state probability of a contact, respectively, $\pi(0) = (1-\alpha)/(1-\alpha+\beta)$ and $\pi(1) = \beta/(1-\alpha+\beta)$.

A mobile user subscribes to receive content updates from publishers. This content is transmitted to users through messages sent by the service provider. The age of a message



(a) (b)

Figure 1. Contact opportunity CDF, (a) probability of a contact followed by no contact and (b) probability of no contact followed by no contact.

is defined as the duration of time (in time slots) since the message was updated the last time. The objective of a user is to minimize the age of its message taking into account its utility and energy and monetary costs. At each time slot the user has to choose between 1) being inactive during the slot 2) sensing and using an access point if available, otherwise waiting for the rest of the slot or 3) sensing and using an access point if available, otherwise using 3G.

The system state at time t is unknown to the user if he does not sense the channel at that time, *i.e.*, $s(t)$ is not fully observable. Let $\lambda(t)$ be the probability that the user has a useful contact opportunity with an AP at the beginning of slot t, given the mobile decision at time $t-1$ and its observation history. $\lambda(t)$ is also referred to as the *belief* of the user about the availability of access points at time t, immediately before the transition from $s(t-1)$ to $s(t)$.

Users decisions are based on two pieces of information: 1) the partial information about the access point state and 2) complete information about the age of the message it owns, which naturally yields a partially observable Markov decision process (POMDP). Assuming that users have complete information about the age of the message that they own corresponds to considering a system at which new updates are available with high probability at every time slot (therefore, the message aging process is deterministic given that users remain inactive). Alternatively, we could easily adapt our model to account for the light load regime, wherein messages do not necessarily age at every time slot even if users remain inactive. In the latter case, users would have incomplete information about the availability of access points as well as about their message ages. Nonetheless, to simplify notation, in the remainder of this paper we assume complete knowledge about age information.

Let x_t be the age of the message owned by a user at time t. Let $\lambda(t)$ be belief of the user about the availability of an access point at time t, immediately before the transition from $s(t-1)$ to $s(t)$. At each time slot t, the user chooses its action $a(t)$,

$$a(t) = \begin{cases} 0, & \text{wait for the next slot;} \\ 1, & \text{sense and use WiFi if available;} \\ 2, & \text{sense and use WiFi if available, otherwise 3G.} \end{cases}$$

We assume that actions are determined at the beginning of each time slot. When the mobile user decides to sense (*i.e.* $a(t) \neq 0$), he observes the availability of access points. We denote by $\theta(t)$ the observation outcome, where $\theta(t) = 1$ if an access point is available, and $\theta(t) = 0$ otherwise. The age of the message increases by one if the user chooses to stay inactive or if he is not in range of an access point, and is reset to one otherwise. Let M be the maximum age of a message. Then,

$$x_{t+1} = \begin{cases} \min(x_t + 1, M), & \text{if } a(t) = 0 \text{ or } [a(t) = 1 \text{ and } \theta(t) = 0]; \\ 1, & \text{if } a(t) = 2 \text{ or } [a(t) = 1 \text{ and } \theta(t) = 1]. \end{cases}$$

At time slot t, the mobile user receives an instantaneous reward $r_t((\lambda, x), a)$ as a result of choosing action a when the system is at state (λ, x). The instantaneous reward consists of two components. A positive component represents the utility of the message the user is willing to update, $U(x)$. We assume that $U(x)$ is a non-increasing function of x, and consider a linear utility defined by $U(x) = M - x$ when deriving closed form expressions. The second component of the instantaneous reward is negative, and corresponds to the consumed energy and the monetary cost for each message transmitted. Let E be the cost incurred to maintain the mobile device active, measured in monetary units. Then, the energy cost e_t is given by

$$e_t(a(t)) = \begin{cases} E, & \text{if } a(t) = 1 \text{ or } a(t) = 2; \\ 0, & \text{if } a(t) = 0. \end{cases}$$

In addition to the cost incurred to maintain the device active and to sense the channel, there is an energy cost incurred to transmit WiFi and 3G packets. Each message transmitted incurs costs C and C_{3G} when transmitted through WiFi and 3G, respectively. Messages might also be associated to monetary charges set by the service provider. The prices

charged to use WiFi and 3G can be taken into account in C and C_{3G}, respectively. When a user receives a message update, he is subject to a cost m_t,

$$m_t(a(t),\theta(t))=\begin{cases} C, & \text{if } [a(t)=1 \text{ or } a(t)=2] \text{ and } \theta(t)=1; \\ C_{3G}, & \text{if } a(t)=2 \text{ and } \theta(t)=0. \end{cases}$$

Monetary and energy costs incurred to transmit content through WiFi are usually lower than the ones to transmit through 3G. Therefore, throughout this paper we assume $C_{3G} > C$. The instantaneous user reward at time t, $r_t((\lambda,x),a)$, is

$$r_t((\lambda,x),a)= \begin{cases} U(x), & \text{if } a(t)=0; \\ U(x)-E, & \text{if } a(t)=1 \text{ and } \theta(t)=0; \\ U(x)-E-C, & \text{if } [a(t)=1 \text{ or } a(t)=2] \text{ and } \theta(t)=1; \\ U(x)-E-C_{3G}, & \text{if } a(t)=2 \text{ and } \theta(t)=0. \end{cases}$$

Assumption 1 (Active at maximum age)**.** *The user chooses to sense the channel if the age of the message reaches its maximum value.*

Assumption 1 is a natural assumption, and corresponds to the fact that when $x_t = M$ the user has no incentive for idle waiting, and will choose between using WiFi if available (action 1) or using 3G otherwise (action 2).

IV. Partially Observable Markov Decision Process Framework

We now describe the partially observable Markov decision process (POMDP) used to derive the optimal sensing policy. One of the key ingredients of the POMDP is the update rule for the belief state of the mobile at time t, $\lambda(t)$. The belief is updated at the end of each time slot based on the action $a(t)$ and the observation outcome $\theta(t)$.

Let $\Omega(.|a(t),\theta(t))$ be the update rule operator on the belief value.[2] The update rule is given by

$$\lambda(t+1)=\Omega(\lambda(t)|a(t),\theta(t)) \tag{1}$$

where

$$\Omega(\lambda(t)|a(t),\theta(t))= \begin{cases} \alpha, & \text{if } [a(t)=1 \text{ or } a(t)=2] \text{ and } \theta(t)=1; \\ \beta, & \text{if } [a(t)=1 \text{ or } a(t)=2] \text{ and } \theta(t)=0; \\ \lambda(t)(\alpha-\beta)+\beta, & \text{if } a(t)=0. \end{cases} \tag{2}$$

A sensing policy μ for our POMDP is given by a vector $[\mu_1,\mu_2,\ldots]$, where each μ_t is a mapping from a system state $(\lambda(t),x_t)$ to an action $a(t)$ to be taken in slot t. We say that a policy is *stationary* if the function μ_t does not depend on time t, but only on the system state. A policy strikes a balance between instantaneous reward gains and information collection for future use. In this paper we restrict to Markovian stationary policies, and in the remainder of this paper we omit the time index t from all variables. It can be shown that the restriction to Markovian stationary policies can be made without loss of generality [14].

We look for an optimal policy that maps each system state (λ,x) to an action a. Our aim is to maximize the expected average reward. We denote the optimal policy by $\mu^\star$. The expected average reward is given by

$$R(r,\mu)=\lim_{T\to\infty}\frac{1}{T}\mathbb{E}_\mu\left[\sum_{t=1}^{T} r_t((\lambda,x),a,\theta)\Big|\lambda(1)\right], \tag{3}$$

where $\mathbb{E}_\mu$ represents the conditional expectation given that policy μ is employed and $\lambda(1)$ is the initial state belief, which can be an arbitrary distribution. The optimal policy $\mu^\star$ is

$$\mu^\star=\arg\max_{\mu}\Big\{R(r,\mu)\Big\}. \tag{4}$$

Let $\mathcal{P}_\mu$ be the transition probability matrix of the Markov chain $\{(x_t,\lambda_t), t=1,2,\ldots\}$ which characterizes the dynamics of the age and of the belief about contact opportunities, for a given policy μ. Let the POMDP *average reward per time slot in steady state*, when policy μ is adopted, be g_μ [14]. As the system transition matrix $\mathcal{P}_\mu$ comprises a single connected component, g_μ is constant and does not depend on the initial system state.

Let $V((\lambda,x),t;\mu)$ be the difference between the expected total reward accumulated by time t when the system starts in (λ,x) and the expected total reward accumulated when the system starts in steady state. Let $r_\mu(\lambda,x)$ be the expected instantaneous reward received in a time slot when the system is at state (λ,x) and policy μ is used; $\mathbf{r}_\mu$ is the vector of expected instantaneous rewards. Let $V((\lambda,x);\mu)$ be the *value function* at state (λ,x), defined as a function of $V((\lambda,x),t;\mu)$ as

$$V((\lambda,x);\mu) = \left(\lim_{l\to\infty} V((\lambda,x),l;\mu)\right)(\lambda,x) \tag{5}$$

$$= \left(\lim_{l\to\infty}\sum_{t=0}^{l}\mathcal{P}_\mu^t(\mathbf{r}_\mu-g_\mu\mathbf{e})\right)(\lambda,x) \tag{6}$$

where $\mathbf{e}$ is a column vector with all its elements equal to one.

Let $Q_a(\lambda,x;\mu)$ be the relative expected average reward obtained by a user that takes action a when the belief value is λ and the age is x. Next, we provide expressions for the relative reward $Q_a(\lambda,x;\mu)$ as a function of the chosen action a.

Case 1 ($a=0$): The user decides to turn off his mobile device, *i.e.*, $a=0$, obtains an instantaneous reward of $U(x)$, the age of the message increases by one, and the belief is updated,

$$Q_0(\lambda,x;\mu)=U(x)+V(\Omega(\lambda|0),x+1;\mu). \tag{7}$$

Case 2 ($a=1$): The user decides to sense, consumes energy E and gets an observation θ. According to this observation, the user turns off his mobile device if $\theta=0$ (in which case the age will increase), or uses WiFi if $\theta=1$ (and the age is reset to one),

$$Q_1(\lambda,x;\mu)=U(x)-E+(1-\lambda)V(\Omega(\lambda|1,0),x+1;\mu) \\ +\lambda[-C+V(\Omega(\lambda|1,1),1;\mu)]. \tag{8}$$

[2] It has been proved in [22] that the belief $\lambda(t)$ together with the update rule operator $\Omega(.|a(t),\theta(t))$ yield a sufficient statistic for the optimal action at time slot t.

Case 3 ($a = 2$): The user decides to sense, and uses 3G if WiFi is not available,

$$Q_2(\lambda, x; \mu) = U(x) - E + (1-\lambda)[-C_{3G} + V(\Omega(\lambda|2,0), 1; \mu)] + \\ + \lambda[-C + V(\Omega(\lambda|2,1), 1; \mu)]. \quad (9)$$

where $\Omega(\lambda|1,1) = \Omega(\lambda|2,1) = \alpha$ and $\Omega(\lambda|1,0) = \Omega(\lambda|2,0) = \beta$.

It follows from [14, eq. (8.4.2)] that a policy $\mu^\star$ which satisfies the following conditions is optimal,

$$g_{\mu^\star} + V(\lambda, x; \mu^\star) = \max_a \Big\{ Q_a(\lambda, x; \mu^\star) \Big\}. \quad (10)$$

The optimal action at state (λ, x) is

$$a^\star(\lambda, x) = \arg\max_a \Big\{ Q_a(\lambda, x; \mu^\star) \Big\}. \quad (11)$$

Henceforth, we also consider the following assumption.

Assumption 2. *Assume that $\alpha \geq \beta$.*

We are currently investigating the extent to which our results still hold if Assumption 2 is not satisfied.

V. Optimal Policy

In this section we present our key results concerning the structure of the optimal policy. After introducing monotonicity properties of the value function in Section V-A, we establish properties about the optimal policy in Section V-B.

A. Monotonicity of Value Function

A first step in establishing the structure of optimal policies is to study the monotonicity of the value function, which by itself can provide insights about the problem under study [12]. The proofs of the results in this section, when not provided in the appendices, are available in a companion technical report [16].

Proposition 1. *The value function $V(\lambda, x)$ is monotonically decreasing with respect to the age of message x, for all belief λ,*

$$V(\lambda, x) > V(\lambda, x'), \text{ for } x < x', 0 \leq \lambda \leq 1 \quad (12)$$

Proposition 1 states that, for a given belief state, an updated message yields higher expected reward than an old message.

Proposition 2. *The value function $V(\lambda, x)$ is monotonically increasing with respect to the belief λ, for all ages x,*

$$V(\lambda, x) > V(\lambda', x), \text{ for } \lambda' < \lambda, x = 1, 2, \ldots, M \quad (13)$$

According to Proposition 2, for any given age, the higher the belief that a contact will occur, the higher the expected reward.

B. Optimal Policy

In this subsection we derive the characteristics of an optimal policy for a mobile user. Intuitively, when the user belief λ about a future contact opportunity is small, and the age of the message x is also small, it is beneficial to remain inactive, and to wait to sense the channel at a future point in time. After sensing the channel, depending on the availability of an access point, the user will decide to get an update. The more outdated is the message, the more likely it is that the user will get updates using the 3G network. In what follows, we formalize the above intuition.

We start by deriving the optimal threshold for the scenario at which a user remains inactive, *i.e.*, when $a^\star(\lambda, x) = 0$,

Proposition 3. *For every information state (λ, x), the user chooses to turn off his mobile device if $\lambda \leq \lambda^\star$ where $\lambda^\star$ satisfies*

$$\lambda^\star = \max\Big\{ \min\{\rho_1(\lambda^\star, x), \rho_2(\lambda^\star, x)\}, 0 \Big\}. \quad (14)$$

and

$$\rho_1(\lambda^\star, x) = \frac{V(\Omega(\lambda^\star|0), x+1) - V(\beta, x+1) + E}{-C + V(\alpha, 1) - V(\beta, x+1)} \quad (15)$$

$$\rho_2(\lambda^\star, x) = \frac{V(\Omega(\lambda^\star|0), x+1) - V(\beta, 1) + E + C_{3G}}{V(\alpha, 1) - V(\beta, 1) - C + C_{3G}} \quad (16)$$

Note that if $-C + V(\alpha, 1) - V(\beta, x+1) = 0$, we have $\lambda^\star = \max(\rho_2(\lambda^\star, x), 0)$. Proposition 3 yields a condition under which a user must stay inactive, as a function of his belief about the probability of future contact opportunities λ, and the message age x. A condition under which a user must become active is determined by the following proposition.

Proposition 4. *For every information state (λ, x), if $\lambda > \pi(1)$ the user must turn on his mobile device.*

According to Proposition 4, if the belief about the probability of a contact opportunity is greater than the stationary probability $\pi(1)$, the user must activate his mobile device.

Proposition 5. *When the user chooses to sense the channel in search for an access point, if WiFi is not available the 3G connection is used if*

$$V(\beta, 1) > V(\beta, x+1) + C_{3G} \quad (17)$$

Note that, according to Proposition 5, the decision of using 3G is independent of the belief about the probability of a contact opportunity, and depends only on the age x and the 3G cost C_{3G}.

VI. Numerical results

We now illustrate the applicability of the proposed model through a set of numerical examples. Our goal is to investigate how the various system parameters influence the optimal policy. To this aim, we consider the following system parameters as our reference setting: the maximum age of a message is $M = 12$, the energy cost is $E = 5$, and the costs for using WiFi and 3G are $C = 10$ and $C_{3G} = 300$, respectively.
We then vary the parameters, considering the following three scenarios.

- Scenario 1: Access points are most of the time available, $\alpha = 0.8$ and $\beta = 0.3$.
- Scenario 2: Access points are often unavailable, $\alpha = 0.2$ and $\beta = 0.6$.
- Scenario 3: Access points are most of the time available and the cost of 3G is low compared to the previous scenarios, $\alpha = 0.8$, $\beta = 0.3$ and $C_{3G} = 100$.

Figures 2, 3 and 4 show the optimal policies in the three scenarios considered above, obtained using value iteration [14], [16]. For each belief λ and age x, a policy determines the probability of adopting each of the available actions. In the three scenarios considered, given a message age x, the optimal policies state that a user must turn off his mobile device and wait for the next slot if $\lambda < \lambda^*$, in accordance to Proposition 3. Once the threshold is exceeded, the user must sense for an access point looking for a useful contact opportunity.

As the message age increases, the user is less likely to remain inactive. In addition, the role of the belief λ decreases as the age x increases. In Figures 2, 3 and 4, at ages 11, 9 and 7, respectively, the user decides to sense and transmit using the 3G if the access point is not available (see Proposition 1). Thus, the message age will never surpass these maximum values at the three considered scenarios.

As the access points become less available, is it beneficial to wait for shorter periods of time before activating the mobile devices, as indicated by the increased set of parameters for which action 1 is selected when switching from scenario 1 to scenario 2. Finally, in scenario 3 the user is more motivated to use 3G as the 3G cost is reduced.

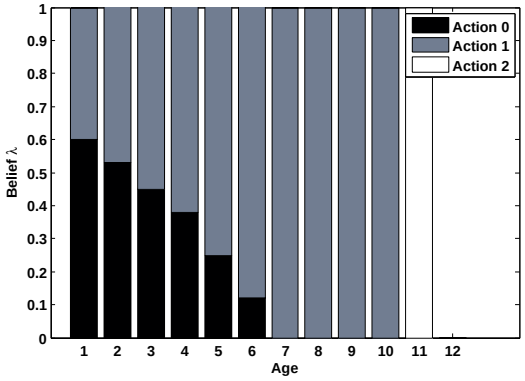


Figure 2. Optimal policy for a mobile user where $\alpha = 0.8$, $\beta = 0.3$ and $C_{3G} = 300$

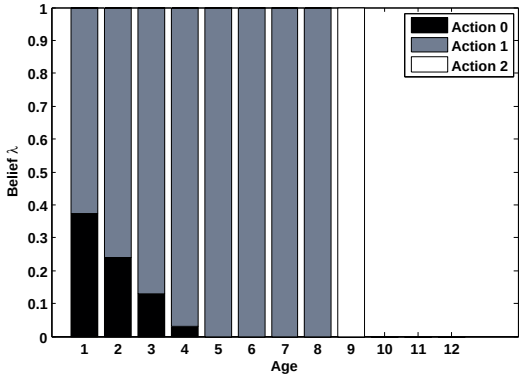


Figure 3. Optimal policy for a mobile user where $\alpha = 0.2$, $\beta = 0.6$ and $C_{3G} = 300$

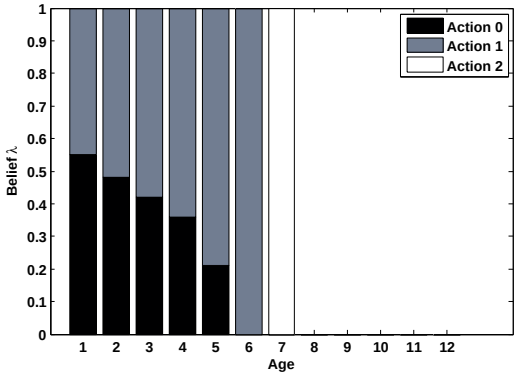


Figure 4. Optimal policy for a mobile user where $\alpha = 0.8$, $\beta = 0.3$ and $C_{3G} = 100$

VII. Related Work and Discussion

There is a vast literature on learning in partially observable markovian decision processes [19], [20], and in particular on learning sensing policies in the realm of cognitive radio networks [22], [11], [23]. Nonetheless, to the best of our knowledge none of the previous works considered the aging control problem under an incomplete information framework.

Focusing on cognitive radio networks, Zhao *et al.* [22] studied decentralized MAC protocols, where secondary users search for spectrum opportunities without a central controller. They propose an analytical framework based on a POMDP, and look for optimal sensing and channel selection schemes that maximize the expected total number of bits delivered over a finite number of slots. Optimal policies were derived depending on the history of decisions and observations of primary channels occupancy. Liu *et al.* [11] consider the interaction between secondary users who are trying to maximize their throughput. They also propose a POMDP in order to find an optimal opportunistic spectrum access policy. In this paper, in contrast, we consider the single controller case, and account for the tradeoff between message aging and power consumption.

Taking into account energy constraints in defining the optimal sensing policies is required to satisfy some QoS requirements. Few works have included the energy consumption in their policies definitions. Chen *et al.* [8] formulated the problem as a POMDP with a finite randomized horizon and Wang *et al.* [21] proposed an adaptive algorithm to find the optimal contention probability that minimizes the expected delay, in the context of a queueing analysis of a cognitive radio network with multiple secondary users. In a queuing context too, Altman *et al.* [1] propose a model based on MDP, where each user chooses dynamically both the power and the admission control to be adopted so as to maximize its expected throughput. The authors then studied the equilibria for the multi-player scenario in a stochastic game context.

Previous work that accounted for the aging control considered the problem from the perspective of providers or publishers [7], [3], [17]. In this paper, we consider the problem from the perspective of users. In [7], using a spatial mean field

approach, the authors model the distribution of message ages in a mobile network. Activation of mobile devices strategies were proposed in [9], [15]. In [15], the authors propose a joint activation and link selection control policy to minimize the energy consumption under delay constraints.

Altman *et al.* [2] consider publishers of evolving files, with the goal of reducing the energy expenditure by controlling the probability of transmitting messages to users. In [2], a Markov Decision Process (MDP) model was proposed to derive the structure of the optimal aging control policy. However, the authors assume that the probability to find a useful contact opportunity between a user and a WiFi access point is constant and independent across time slots. This is a strong assumption, since the correlations between contact opportunities experienced by a user are present in real mobile network. In this paper, we overcome this limitation by using a POMDP rather than a MDP. We study the monotonicity of the value function to establish the structure of the optimal policies. The activation policy of mobile users in our model depends not only on the messages age, as in [2], but also on the belief about the probability of finding a useful contact opportunity.

VIII. Conclusion and perspectives

In this paper, we have developed a POMDP framework to study aging control in hybrid wireless networks, taking into account the tradeoff between energy consumption and the age of messages. From the DieselNet measurements, we learned that contact opportunities between users and APs are unpredictable but correlated. We then derived a sensing policy for the aging control problem, and we have shown several properties satisfied by optimal policies.

We believe this work opens several avenues for future research. While in this paper we have studied the aging control problem from the perspective of one single user, future work consists of considering the interaction between several mobile users. Another direction is to consider strategic service providers, that adjust their prices accounting for users that adopt sensing policies described in this paper.

References

[1] E. Altman, K. Avrachenkov, N. Bonneau, M. Debbah, R. El-Azouzi, and D. Menasche. Constrained stochastic games in wireless networks. In *GLOBECOM,315-320*, 2007.

[2] E. Altman, R. El-Azouzi, D. Menasche, and Y. Xu. Forever young: Aging control for smartphones in hybrid networks. In *IFIP Performance (Poster Session)*, 2011.

[3] E. Altman, P. Nain, and J. Bermond. Distributed storage management of evolving files in delay tolerant ad hoc networks. In *INFOCOM,1431-1439*, 2009.

[4] A. Balasubramanian, R. Mahajan, and A. Venkataramani. Augmenting mobile 3g using wifi. In *MobiSys,209-222*, 2010.

[5] A. Balasubramanianm, B. N. Levine, and A. Venkataramani. Enabling interactive applications for hybrid networks. In *MobiCom,70-80*, 2008.

[6] A. Cassandra, L. Kaelbling, and M. Littman. Acting optimally in partially observable stochastic domains. In *Proc. Conf. on Artificial Intelligence (AAAI)*, Seattle, 1994.

[7] A. Chaintreau, J.-Y. L. Boudec, and N. Ristanovic. The age of gossip: spatial mean field regime. In *SIGMETRICS,109-120*, 2009.

[8] Y. Chen, Q. Zhao, and A. Swami. Distributed spectrum sensing and access in cognitive radio networks with energy constraint. In *IEEE Transactions on Signal Processing*, 2009.

[9] L. B. Le, E. Modiano, and N. B. Shroff. Optimal control of wireless networks with finite buffers. In *INFOCOM,2034-2042*, 2010.

[10] M. L. Littman, A. R. Cassandra, and L. P. Kaelbling. Learning policies for partially observable environments: Scaling up. In *Proceedings of 12th International Conference on Machine Learning*, 1995.

[11] H. Liu, B. Krishnamachari, and Q. Zhao. Cooperation and learning in multiuser oppoertunistic spectrum access. In *IEEE ICC*, 2008.

[12] W. S. Lovejoy. Some monotonicity results for partially observed Markov decision processes. *Operations Research*, 35(5):736–743, 1987.

[13] G. Monahan. A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science*, 28(1):1–16, 1982.

[14] M. L. Puterman. *Markov Decision Process Discrete Stochastic Dynamic Programming*. Wiley Series in Probability and Statistics, 2005.

[15] M. R. Ra, J. Paek, A. B. Sharma, R. Govindan, M. H. Krieger, and M. J. Neely. Energy-delay tradeoffs in smartphone applications. In *MobiSys,255-270*, 2010.

[16] M. Raiss-El-Fenni, R. El-Azouzi, D. Menasche, and Y. Xu. Optimal sensing policies for smartphones in hybrid networks: A POMDP approach. In *Technical report*, 2012.

[17] J. Reich and A. Chaintreau. The age of impatience: optimal replication schemes for opportunistic networks. In *CONEXT,85-96*, 2009.

[18] C. Sherlaw-Johnson, S. Gallivan, and J. Burridge. Estimating a Markov transition matrix from observational data. *Journal of the Operational Research Society*, 46(3):405–410, 1995.

[19] S. Singh, T. Jaakkola, and M. Jordan. Learning without state estimation in partially observable markovian decision processes. In *ICML,284-292*, 1994.

[20] S. Thrun. Monte Carlo POMDPs. In S. Solla, T. Leen, and K.-R. Muller, editors, *Advances in Neural Information Processing Systems*, pages 1064–1070. MIT Press, 2000.

[21] S. Wang, J. Zhang, and L. Tong. Delay analysis for cognitive radio networks with random access: A fluid queue view. In *INFOCOM,1055-1063*, 2010.

[22] Q. Zhao, L. Tong, A. Swami, and Y. Chen. Decentralized cognitive mac for opportunistic spectrum access in ad hoc networks: A POMDP framework. *IEEE JSAC*, 25(3), April 2007.

[23] H. Zheng and C. Peng. Collaboration and fairness in opportunistic spectrum access. In *IEEE ICC,3132-3136*, 2005.

Appendix

Proof of proposition 1

Next, we prove that the value function $V(\lambda, x)$ is monotonically decreasing with the age of message x for all belief values λ. According to Assumption 1, when the age reaches the maximum value $x = M$, the user chooses to sense ($a = 1$ or $a = 2$). We consider two cases,

Case 1) $Q_2(\lambda, M) > Q_1(\lambda, M)$

At age $x = M$, it follows from (10) that

$$\begin{aligned} & g_\mu + V(\lambda, M) = && (18)\\ = \;& Q_2(\lambda, M) && (19)\\ = \;& U(M) - E + (1-\lambda)(-C_{3G} + V(\beta, 1)) + \lambda(-C + V(\alpha, 1)) \end{aligned}$$

At age $x = M - 1$, it follows from (10) that

$$\begin{aligned} & g_\mu + V(\lambda, M-1) = && (20)\\ = \;& \max\left\{Q_0(\lambda, M-1); Q_1(\lambda, M-1); Q_2(\lambda, M-1)\right\} \\ \geq \;& Q_2(\lambda, M-1) && (21)\\ = \;& U(M-1) - E + \\ & \quad (1-\lambda)(-C_{3G} + V(\beta, 1)) + \lambda(-C + V(\alpha, 1)) && (22)\\ \geq \;& U(M) - E + \\ & \quad (1-\lambda)(-C_{3G} + V(\beta, 1)) + \lambda(-C + V(\alpha, 1)) && (23)\\ = \;& Q_2(\lambda, M) = g_\mu + V(\lambda, M) && (24) \end{aligned}$$

Therefore, $V(\lambda, M-1) \geq V(\lambda, M)$, and the result is proved for $x = M-1$. To prove the result for $x < M-1$, we use backward induction,

Initial condition: $V(\lambda, M-1) \geq V(\lambda, M), \quad 0 \leq \lambda \leq 1$

Induction hypothesis: Given x_0, $1 < x_0 \leq M-1$, we assume that $V(\lambda, x) \geq V(\lambda, x+1), \quad x_0 \leq x \leq M-1, \quad 0 \leq \lambda \leq 1$.

Induction step: Next, we show that if the result holds for x it holds for $x-1$, $x > 1$,

$$
\begin{aligned}
&V(\lambda, x-1) = \\
&= \max\Big\{Q_0(\lambda, x-1); Q_1(\lambda, x-1); Q_2(\lambda, x-1)\Big\} && (25)\\
&= U(x-1) + \max\Big\{V(\Omega(\lambda|0), x); && (26)\\
&\quad -E + (1-\lambda)V(\beta, x) + \lambda(-C + V(\alpha, 1)); \\
&\quad -E + (1-\lambda)(-C_{3G} + V(\beta, 1)) + \lambda(-C + V(\alpha, 1))\Big\} \\
&\geq U(x) + \max\Big\{V(\Omega(\lambda|0), x+1); && (27)\\
&\quad -E + (1-\lambda)V(\beta, x+1) + \lambda(-C + V(\alpha, 1)); \\
&\quad -E + (1-\lambda)(-C_{3G} + V(\beta, 1)) + \lambda(-C + V(\alpha, 1))\Big\} \\
&\geq V(\lambda, x)
\end{aligned}
$$

where (27) follows from (26) due to the induction hypothesis.

Case 2) $Q_1(\lambda, M) \geq Q_2(\lambda, M)$

The proof for case 2) is similar to that for case 1). ■

Proof of proposition 2

We now prove that the value function $V(\lambda, x)$ is monotonically increasing with respect to the belief λ, $x = 0, 1, \ldots, M$. According to Assumption 1, when the age reaches the maximum value $x = M$, the user chooses to sense ($a = 1$ or $a = 2$). We consider two cases,

Case 1) We assume that

$$Q_1(\lambda, M) \geq Q_2(\lambda, M) \tag{28}$$

In this case, we rely on the following property, whose proof is provided in Appendix A.

$$-C + V(\alpha, 1) \geq V(\beta, M) \tag{29}$$

Given a real number λ', $0 < \lambda' < 1$, for all $\lambda \leq \lambda'$ we have

$$
\begin{aligned}
&V(\lambda, M) = && (30)\\
&= -g_\mu + U(M) - E + \\
&\quad (1-\lambda)V(\beta, M) + \lambda(-C + V(\alpha, 1)) && (31)\\
&= -g_\mu + U(M) - E + \\
&\quad V(\beta, M) + \lambda(-C + V(\alpha, 1) - V(\beta, M)) && (32)\\
&\leq -g_\mu + U(M) - E + \\
&\quad V(\beta, M) + \lambda'(-C + V(\alpha, 1) - V(\beta, M)) && (33)\\
&= V(\lambda', M) && (34)
\end{aligned}
$$

where (33) follows from (32) due to (29).

Using backward induction, we have

Initial condition: $V(\lambda, M) \leq V(\lambda', M), \quad 0 \leq \lambda \leq \lambda'$

Induction hypothesis: Given x_0, $1 < x_0 \leq M-1$, we assume that $V(\lambda, x+1) \leq V(\lambda', x+1), \quad x_0 \leq x \leq M-1, \quad 0 \leq \lambda \leq 1$.

Induction step: Next, we show that if the result holds for $x+1$ it holds for x, $x \geq 1$.

Recall that $\pi(1)$ is the stationary probability that the access point is available. Then, $\Omega(\pi(1)) = \pi(1)$, and $\pi(1) = \beta/(1-\alpha+\beta)$. In Appendix A we show that

$$\lambda \leq \pi(1) \Rightarrow \Omega(\lambda) \geq \lambda \tag{35}$$

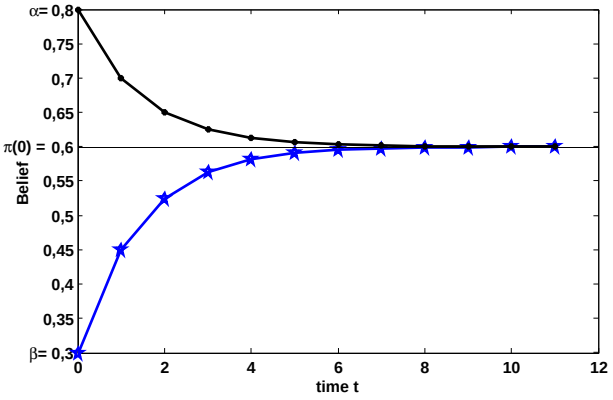


Figure 5. Evolution of $\lambda(t)$ for $a = 0$ ($\alpha = 0.8$, $\beta = 0.3$).

Figure 5 illustrates the evolution of the belief λ over time, obtained from (2), for $a = 0$, when $\alpha \geq \beta$ (see Assumption 2).

First, we show that $Q_0(\lambda, x) \leq Q_0(\lambda', x)$.

$$
\begin{aligned}
Q_0(\lambda, x) &= U(x) + V(\Omega(\lambda|0), x+1) && (36)\\
&\leq U(x) + V(\Omega(\lambda'|0), x+1) && (37)\\
&= Q_0(\lambda', x)
\end{aligned}
$$

Second, we show that $Q_2(\lambda, x) \leq Q_2(\lambda', x)$.

$$
\begin{aligned}
Q_2(\lambda, x) &= U(x) - E + \lambda(-C + V(\alpha, 1)) && (38)\\
&\quad +(1-\lambda)(-C_{3G} + V(\beta, 1)) \\
&= U(x) - E - C_{3G} + V(\beta, 1) && (39)\\
&\quad +\lambda(-C + V(\alpha, 1) + C_{3G} - V(\beta, 1))
\end{aligned}
$$

As $Q_1(\lambda, M) \geq Q_2(\lambda, M)$ (see (28)), it follows that

$$V(\beta, M) > -C_{3G} + V(\beta, 1) \tag{40}$$

(40) together with (29) yields

$$-C + V(\alpha, 1) + C_{3G} - V(\beta, 1) > 0 \tag{41}$$

Replacing (41) into (39),

$$
\begin{aligned}
Q_2(\lambda, x) &\leq U(x) - E - C_{3G} + V(\beta, 1) \\
&\quad +\lambda'(-C + V(\alpha, 1) + C_{3G} - V(\beta, 1)) \\
&\leq Q_2(\lambda', x) && (42)
\end{aligned}
$$

Third, we show that $Q_1(\lambda, x) \leq Q_1(\lambda', x)$.

If $-C+V(\alpha,1)-V(\beta,x+1)\geq 0$,

$$\begin{aligned} Q_1(\lambda,x) &= U(x)-E+V(\beta,x+1) && (43)\\ &\quad +\lambda(-C+V(\alpha,1)-V(\beta,x+1)) \\ &\leq U(x)-E+V(\beta,x+1) && (44)\\ &\quad +\lambda'(-C+V(\alpha,1)-V(\beta,x+1)) \\ &\leq Q_1(\lambda',x) && (45)\end{aligned}$$

If $-C+V(\alpha,1)-V(\beta,x+1)<0$,

$$\begin{aligned} Q_1(\lambda,x) &\leq U(x)-E+V(\beta,x+1) && (46)\\ &\leq U(x)-E+V(\Omega(\lambda|0),x+1) && (47)\\ &\leq Q_0(\lambda,x) && (48)\end{aligned}$$

$Q_1(\lambda,x)\leq Q_0(\lambda,x)$ implies that the value function can either be equal to $Q_0(\lambda,x)$ or $Q_2(\lambda,x)$. Then

$$g_\mu+V(\lambda,x)=\max\Big\{Q_0(\lambda,x),Q_2(\lambda,x)\Big\}.$$

We have shown that $Q_0(\lambda,x)$, $Q_1(\lambda,x)$ and $Q_2(\lambda,x)$ are increasing with respect to the belief. Thus $V(\lambda,x)\leq V(\lambda',x)$ for $\lambda\leq\lambda'$ $(x=0,1,\ldots,M)$.

Case 2) We assume that $Q_2(\lambda,M)>Q_1(\lambda,M)$. In this case, we can show that

$$-C+V(\alpha,1)\geq -C_{3G}+V(\beta,1) \quad (49)$$

The proof of (49) is provided in Appendix A. Using (49), the remainder of the proof for case 2) is similar to that for case 1). ■

Proof of (35) **:**

As we assumed that $\alpha\geq\beta$, the update function $\Omega(\lambda)$ is increasing with respect to the belief λ. We now show by induction on time t that if $\lambda(t_0)\leq\pi(1)$ then $\Omega(\lambda(t))\geq\lambda(t)$ for $t\geq t_0$.

We assume, without loss of generality, that $\lambda(t_0)=\beta$.

Initial condition: Note that $\beta\leq\pi(1)$ and $\Omega(\beta)=(\alpha-\beta)\beta+\beta\geq\beta$

Induction hypothesis: Assume that if $\lambda(t)\leq\pi(1)$ then $\Omega(\lambda(t))\geq\lambda(t)$, $t_0\leq t\leq t_1$.

Induction step: We want to show that if $\lambda(t+1)\leq\pi(1)$ then $\Omega(\lambda(t+1))\geq\lambda(t+1)$,

$$\begin{aligned} \Omega(\lambda(t+1)) &= \Omega(\Omega(\lambda(t))) && (50)\\ &= (\alpha-\beta)\Omega(\lambda(t))+\beta && (51)\\ &\geq (\alpha-\beta)\lambda(t)+\beta && (52)\\ &= \alpha\lambda(t)+\beta(1-\lambda(t)) && (53)\\ &\geq \lambda(t+1) && (54)\end{aligned}$$

Where (52) follows from (51) by the induction hypothesis, and (54) follows from (53) due to (2). ■

Proof of (29) **:**

We prove (29) by contradiction. Suppose that

$$-C+V(\alpha,1)<V(\beta,M) \quad (55)$$

If $\lambda=\beta$ and $x=M$, it follows from (10) that

$$g_\mu = U(M)-E+\beta(-C+V(\alpha,1)-V(\beta,M)) \quad (56)$$

If $\lambda=\alpha$ it follows from (10) that

$$\begin{aligned} g_\mu+V(\alpha,1) &= \max\Big\{Q_1(\alpha,1);Q_2(\alpha,1)\Big\} && (57)\\ &\geq Q_2(\alpha,1) && (58)\\ &\geq U(1)-E+\alpha(-C+V(\alpha,1)) && (59)\\ &\quad +(1-\alpha)(-C+V(\alpha,1)) \\ &\geq U(1)-E-C+V(\alpha,1) && (60)\\ g_\mu &\geq U(1)-E-C\geq 0 && (61)\end{aligned}$$

From (56) and (61) we have

$$\beta(-C+V(\alpha,1)-V(\beta,M))\geq U(1)-C\geq 0 \quad (62)$$

(62) is a contradiction with (55). Therefore. $-C+V(\alpha,1)\geq V(\beta,M)$. ■

Proof of (49) **:**

We prove (49) by contradiction. Suppose that

$$-C+V(\alpha,1)<-C_{3G}+V(\beta,1) \quad (63)$$

Then, it follows from (10) that

$$\begin{aligned} g_\mu+V(\alpha,1) &= \max\Big\{Q_1(\alpha,1);Q_2(\alpha,1)\Big\} \\ &\geq Q_2(\alpha,1) && (64)\\ &\geq U(1)-E+\alpha(-C+V(\alpha,1)) \\ &\quad +(1-\alpha)(-C+V(\alpha,1)) && (65)\\ &\geq U(1)-E-C+V(\alpha,1) && (66)\\ g_\mu &\geq U(1)-E-C\geq 0 && (67)\end{aligned}$$

Given the message age x, let us find the optimal action a.

At $x=M-1$,

$$\begin{aligned} Q_0(\lambda,M-1) &\leq V(\lambda,M)+(\Omega(\lambda|0)-\lambda)\times \\ &\quad (-C+V(\alpha,1)+C_{3G}-V(\beta,1)) && (68)\\ &\leq V(\lambda,M)\leq V(\lambda,M-1) && (69)\end{aligned}$$

It follows from (69) and (67) that

$$Q_0(\lambda,M-1)<V(\lambda,M-1)+g_\mu \quad (70)$$

(70) together (10) imply that action 0 is not the optimal action at $M-1$.

$$\begin{aligned} &Q_1(\lambda,M-1)-Q_2(\lambda,M-1)= \\ &\quad =(1-\lambda)(V(\beta,M)+C_{3G}-V(\beta,1)) && (71)\\ &\quad \leq (1-\lambda)(U(M)-E) && (72)\\ &\quad \leq -E(1-\lambda)\leq 0 && (73)\end{aligned}$$

It follows from (73) and (70) that action 2 is the optimal action at $M-1$.

We proceed with backward induction.

Initial condition: Action 2 is the optimal action at $M-1$.

Induction hypothesis: Given $x_0 > 0$, assume that action 2 is the optimal action at x, $x \geq x_0$.

Induction step: We now show that if action 2 is the optimal action at x, $x \geq x_0$, it is also the best action at $x-1$.

From (7) we have

$$\begin{aligned} Q_0(\lambda, x-1) &\leq V(\lambda, x) + (\Omega(\lambda|0) - \lambda)(-C + V(\alpha, 1) \\ &\quad + C_{3G} - V(\beta, 1)) \quad (74) \\ &\leq V(\lambda, x) \leq V(\lambda, x-1) \quad (75) \end{aligned}$$

Thus, it follows from (75) that action 0 is not the optimal action at $x-1$.

$$\begin{aligned} Q_1(\lambda, x-1) &= U(x-1) - E + \lambda(-C + V(\alpha, 1)) \\ &\quad + (1-\lambda)V(\beta, x) \quad (76) \\ &< U(x-1) - E + \lambda(-C + V(\alpha, 1)) \\ &\quad + (1-\lambda)(-C_{3G} + V(\beta, 1)) \quad (77) \\ &< Q_2(\lambda, x-1) \quad (78) \end{aligned}$$

It follows from (78) and (75) that action 2 is the optimal action at $x-1$, given that it is optimal at x.

Action 2 is the optimal action for $\lambda \leq \pi(1)$ and $x \geq 0$. Considering the special case $\lambda = \beta$, (10) and (9) yield

$$\begin{aligned} V(\beta, 1) &= U(1) - E + \beta(-C + V(\alpha, 1)) \\ &\quad + (1-\beta)(-C_{3G} + V(\beta, 1)) - g_\mu \quad (79) \end{aligned}$$

It follows from (67) and (79) that

$$\beta(-C + V(\alpha, 1) + C_{3G} - V(\beta, 1)) - C_{3G} \geq -C \quad (80)$$

Then, $C_{3G} \geq C$ together with (80) yields a contradiction with (63). Therefore, $-C + V(\alpha, 1) \geq -C_{3G} + V(\beta, 1)$. ■

Proof of proposition 3

At each time slot, the user will decide to idle wait if the expected average reward of this action is larger than the one obtain through other actions.

Action $a(\lambda, x) = 0$ is at least as good as $a(\lambda, x) = 1$ iff

$$\begin{aligned} &U(x) + V(\Omega(\lambda|0), x+1) \geq \\ &\quad (1-\lambda)V(\beta, x+1) + \lambda(-C + V(\alpha, 1)) + U(x) - E \quad (81) \\ &V(\Omega(\lambda|0), x+1) \geq \\ &\quad -E + V(\beta, x+1) + \lambda(-C + V(\alpha, 1) - V(\beta, x+1)) \quad (82) \end{aligned}$$

If $-C + V(\alpha, 1) - V(\beta, x+1) \neq 0$, (15) follows from (82), and action $a(\lambda, x) = 0$ is at least as good as $a(\lambda, x) = 1$ if (82) holds. If $-C + V(\alpha, 1) - V(\beta, x+1) = 0$, (82) together with Proposition 2 also imply that action 0 is at least as good as action 1. Action $a(\lambda, x) = 0$ is at least as good as $a(\lambda, x) = 2$ iff

$$\begin{aligned} &U(x) + V(\Omega(\lambda|0), x+1) \geq \\ &\quad (1-\lambda)(-C_{3G} + V(\beta, 1)) + \lambda(-C + V(\alpha, 1)) + U(x) - E \\ &V(\Omega(\lambda|0), x+1) \geq -E + V(\beta, 1) \\ &\quad -C_{3G} + \lambda(-C + V(\alpha, 1) + C_{3G} - V(\beta, 1)) \quad (83) \end{aligned}$$

As $V(\alpha, 1) > V(\beta, 1)$ and $C_{3G} > C$, we have that $V(\alpha, 1) - V(\beta, 1) - C + C_{3G} > 0$ and the definition of $\rho_2(\lambda, x)$ in (16) follows from (83). ■