Optimal WiFi Sensing via Dynamic Programming

Abhinav Kumar, Sibi Raj B Pillai
Electrical Engineering Department
Indian Institute of Technology Bombay
Mumbai
email: {abhinav.kumar@, bsraj@ee}.iitb.ac.in

Rahul Vaze
School of Technology and Computer Science
Tata Institute of Fundamental Research
Mumbai
email: vaze@tcs.tifr.res.in

Aditya Gopalan
Department of ECE
Indian Institute of Science
Bangalore
e-mail: aditya@ece.iisc.ernet.in

***Abstract*—The problem of finding an optimal sensing schedule for a mobile device that encounters an intermittent WiFi access opportunity is considered. At any given time, the WiFi is in any of the two modes, ON or OFF, and the mobile's incentive is to connect to the WiFi in the ON mode as soon as possible, while spending as little sensing energy. We introduce a dynamic programming framework which enables the characterization of an explicit solution for several models, particularly suitable when the OFF periods are exponentially distributed.**

While the problem for non-exponential OFF periods is ill-posed in general, a usual workaround in literature is to make the mobile device aware if one ON period is completely missed. In this restricted setting, using the DP framework, the deterministic nature of the optimal sensing policy is established, and value iterations are shown to converge to the optimal solution. Finally, we address the blind situation where the distributions of ON and OFF periods are unknown. A continuous bandit based learning algorithm that has vanishing regret (loss compared to the optimal strategy with the knowledge of distributions) is presented, and comparisons with the optimal schemes are provided for exponential ON and OFF periods.

I. Introduction

The available WiFi connectivity in mobile environments can be intermittent. In an effort to maximize WiFi connectivity time, current smartphones keep scanning/sensing for WiFi connection quite frequently, however, they loose precious battery life in this process. The sensing schedule clearly depends on the distributions of the ON and OFF periods of the WiFi access points (AP). This paper is an effort in finding the optimal sensing periods given the knowledge of the ON and OFF period distributions.

Given a geographical area with a fixed number of WiFi APs and a roaming mobile, the WiFi connection opportunity can be modeled as a two-state Markov chain with {ON, OFF} states. This assumption is reasonable for moving users either while driving or walking, where each user moves in an and out of AP coverage areas. In dense areas, the ON-OFF cycles could be periodic, while in semi-urban or rural areas, ON-OFF could follow a certain distribution. In [1], it is shown that the ON and OFF periods can be well approximated by exponential distributions, via analyzing WLAN trace data available from Crawdad [2]. In this paper, we consider general ON-OFF period distributions, and try to find optimal mobile sensing schedules to maximize throughput with minimum sensing cost.

Clearly, there is an inherent tradeoff between sensing cost and WiFi connectivity time: more frequent sensing leads to faster WiFi connectivity however incurs larger cost. Thus, to find the optimal sensing durations, a natural problem is to minimize the sum of the expected length of the missed ON periods and the expected sensing cost as first proposed in [3]. Even though [3] proposed this metric, however, for analysis, the metric was simplified, for example by replacing some of the random variables with their expectations. Optimal solutions to these approximations for general ON and OFF distributions were presented in [3].

In prior work, without explicitly counting for the sensing cost, [1] also found the optimal sensing durations that minimize the rate of missed ON periods. The analysis, however, is not completely rigorous, for example, the missed ON period in a given time period does not depend on the length of the time period, which is anomalous. Some heuristic solutions [1], [4] have also been found that modulate the sensing durations given the frequency of failure of detection. Some other practical smart sensing protocols for WiFi sensing can be found in [5], [6]. Sensing in cognitive radio is also similar to this work [7], however there, the unlicensed users sense to maximize their throughput without harming the licensed users. The cognitive radio setting also leads to a partially observed Markov decision process.

In this paper, we consider the metric as the sum of the expected length of the missed ON periods and the expected sensing cost similar to [3]. Unlike [3], [1], our approach relies on a dynamic programming (DP) formulation. We first solve for the problem when the OFF period is exponentially distributed, while the ON periods are IID with any arbitrary distribution. We give explicit answers in terms of the ON period distribution and the mean of the exponential OFF periods. The DP framework also allows us to rectify the anomalies in the past work concerning exponential ONs and OFFs [1].

For non-exponentially distributed OFF periods, we consider the restriction of one OFF and ON period similar to [3], since otherwise the problem becomes ill-posed, see Remark 1 for a detailed explanation. More importantly, we do not change the metric to suit analysis as done in [3], i.e. the sum of expected length of the missed ON periods and the expected sensing cost is still the metric that we minimize. Again posing the problem as a dynamic program, we obtain structural results to show that the optimal policy is deterministic, and can be found via value iteration that converges to the optimal solution. The restricted

978-3-9018-8273-9/15/ ©2015 IFIP

problem can be seen as a generalization of [8], where the ON period never expires.

Almost all prior work on smart WiFi sensing assumes the knowledge of the distribution of the OFF and ON period distributions. In practice, these can be obtained only via training, which however, is costly in terms of resources. To overcome this, we propose a *blind learning* framework, where the learning algorithm adapts the optimal sensing duration iteratively, without any explicit training. The proposed algorithm is inspired by algorithms for continuous bandit problems [9], [10], where each agent has a continuum of strategies to choose and its objective is to maximize a reward function, however, it does not know the reward distribution conditioned on its choice. We show that the proposed algorithm (following [10]) for finding the optimum sensing durations has a vanishing regret as a function of time, where regret is defined as the difference between the respective rewards of an optimal algorithm with the knowledge of the distribution, and the blind learning algorithm. For ease of exposition, we illustrate the vanishing regret result only when the underlying ON-OFF period distribution is exponential, but as will be readily seen it applies for any other ON-OFF period distributions. To the best of our knowledge this is the first such result in the area of Wi-Fi sensing, and is expected not only to be very useful in practice, but also helpful in the analysis more complex models.

We highlight the main contributions of the paper by listing them below.

- We propose a dynamic programming framework to identify the optimal sensing duration in intermittent WiFi access. The framework applies to arbitrary ON period distributions and exponential OFF period distributions, see Section IV.
- A restricted DP framework gives structural results on optimal policies for the non-exponential OFF periods (Section V).
- We propose a blind learning framework in Section VI, which iteratively identifies the optimal sensing duration, even in the absence of statistical knowledge of the ON-OFF periods.

II. Organization

The rest of the paper is organized as follows. We discuss the system model, together with the problem formulation in Section III. In Section IV, we present our results when the OFF periods are exponentially distributed. The non-exponential OFF periods case is discussed in Section V, where we obtain some structural results. In Section VI, we discuss the learning algorithm which is blind to the ON-OFF period distributions, and is shown to have a vanishing regret. Finally, in Section VII, we present some simulation results for the learning algorithm's regret function, and the achieved WiFi connectivity time.

III. System Model

Consider a mobile device that is moving in and out of WiFi APs' transmission radii, and encounters intermittent WiFi access opportunities in time, as shown in Fig. 1. We assume that at time t, AP state is OFF if the mobile device is not in any AP's transmission radius, and ON otherwise. Thus, as shown in Fig. 1, the mobile sees alternating ON and OFF periods, where it can possibly receive data only during the ON periods.

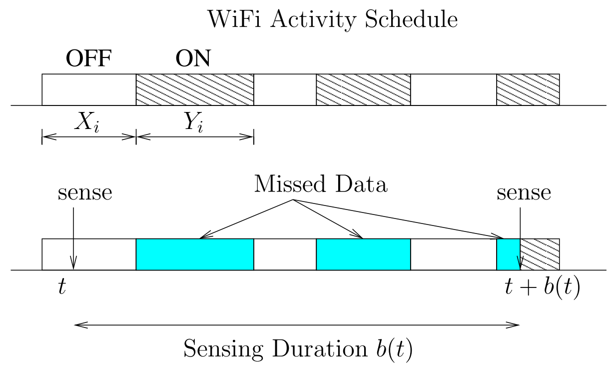

Fig. 1. System model description

To detect ON periods, the mobile device employs sensing. On sensing at time t, if the AP state is found ON, the device gets connected to the AP till the end of that ON period. We assume that the device learns about the disconnection as soon as the ON period is over, by using either the rapid increase in error probability or no useful data transmission. Otherwise, if the AP state is OFF when sensing at time t, then the mobile decides to sleep for the duration till the next sensing epoch $t + b(t)$, as shown in Fig. 1.

To save energy, the mobile senses intermittently, and consequently loses out on connecting to the AP as soon as the ON period starts. In particular, the shaded region (missed data) in Fig. 1 represents the lost opportunity because of intermittent sensing. Longer sleep periods incur less sensing power consumption but decrease the WiFi connectivity time utilized, while shorter sleep periods increase the WiFi connectivity time at the cost of increasing the sensing power consumption.

To strike a balance between the lost ON period time and the sensing power consumption, we consider the problem of finding the sensing intervals so as to minimize the sum of expected lost opportunity for data reception and the expected power for sensing. Let us now make this formal.

Let the duration of the $i^{th}, i \geq 1$ OFF and ON period be denoted by X_i and Y_i, respectively, see Fig. 1. We assume that both X_i and Y_i are independent for $i \geq 1$. The PDF of X and Y is denoted by $f_d(x)$ and $f_c(y)$, where the subscript d and c represent disconnection and connection, respectively.

If a sensing reveals the AP state to be ON, there is no decision to make, and the mobile device stays connected from there till the end of the current ON period, then gets disconnected, and the system restarts. The non-trivial decision problem is encountered when the current sense reveals the AP state to be OFF. We define an ON period to be a *discovered* ON period, if a sensing epoch lies in that ON period. In a discovered ON period, *useful* ON period is the time between the sensing epoch and the end of the discovered ON period. An illustration is provided in Fig. 2. Time period between the end of two consecutive discovered ON periods is defined to be a *session*. Recall that system resets at the the end of each discovered ON period, thus we focus on any one particular session from here onwards.

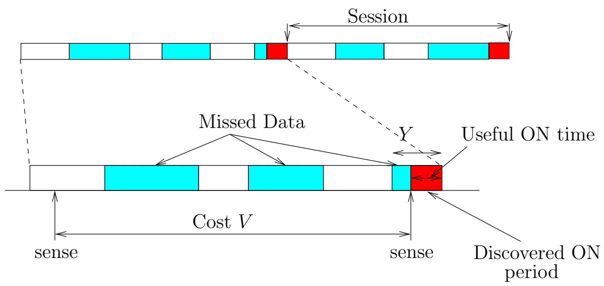

Fig. 2. Illustration of sessions for learning algorithm

Let $\mathbf{1}_{\mathsf{on}}(t)$ ($\mathbf{1}_{\mathsf{off}}(t)$) denote the event that the AP is in ON (OFF) state at time t. Let $P_{\mathsf{off}}(t)$ be the probability (density) that the AP is in OFF state at time t, and $P_{\mathsf{off}}(t+x|t \in *)$ be the probability that the AP is in OFF state at time $t+x$ given that AP is in state $* \in \{\text{OFF, ON}\}$ at time t. We define the cost between two sensing epochs at t and $t+b(t)$ in a session as

$$c(t, b(t)) = c_s + M(t, b(t)), \tag{1}$$

where c_s is a fixed sensing cost, and $M(t, b(t))$ is the missed/lost ON periods (possibly multiple periods are missed because of lack of sensing) between time t and $t+b(t)$. Now, the sensing problem can be cast as a dynamic program (DP),

$$V(t) = \max_{b(t)\geq 0}[\mathbb{E}\{c(t, b(t))\}+P_{\mathsf{off}}(t+b(t)|t \in \mathsf{off})V(t+b(t))], \tag{2}$$

where we have assumed that $b(t)$ is selected as the next sensing duration at time instant t. In this case, the running cost is $\mathbb{E}\{c(t, b(t))\}$, and the process restarts if the AP is in OFF state at time $t + b(t)$, this event happens with probability $P_{\mathsf{off}}(t + b(t)|t \in \mathsf{off})$.

Given that an OFF period is in progress at time t, we define a random variable r_t which represents the residual OFF period (time of the completion of OFF period) at time t. The PDF of r_t is denoted as $f_{r_t}(\cdot)$, and the corresponding CDF is $F_{r_t}(x) = P(X > x|X > t)$, where X represents the duration of the OFF period. Note that $P_{\mathsf{off}}(t + b(t)|t \in \mathsf{off}) = f_{r_t}(b(t))$ in (2).

Remark 1: If the distribution of the OFF periods is not exponential, then (2) is not well-defined. To see this, consider that if for two consecutive sensing times t and $t + b(t)$, the AP state is OFF, the distribution of the residual OFF period starting from $t+b(t)$ ($i.e. r_{t+b(t)}$) is not well-defined since we do not know when the current OFF period started. To handle the non-exponential distribution of the OFF periods, we will follow the approach of [3] in Section V, where it is assumed that as soon as any one complete ON interval is lost/missed because of no sensing epoch lying in that ON period, the system is reset, and the mobile device is notified that an OFF period has begun.

IV. EXP-OFF PERIOD

In light of Remark 1, in this section, we restrict our attention to exponential distribution for the OFF period, while the ON period is allowed to have an arbitrary distribution. Under this assumption, we show that the optimal control $b(t)$ does not depend on t.

Lemma 2: The optimal control $b(t)$ that solves (2) does not depend on t, when OFF periods are exponentially distributed.

Proof: At time t, the next sensing duration $b(t)$ is decided only if $t \in \mathsf{off}$. However, because of the memoryless property of the OFF periods, the event that $t \in \mathsf{off}$ gives no information about the future length of OFF and ON periods, the optimal sensing duration $b(t)$ does not depend on t. ■

With arbitrary ON period distribution, we need the following notation. Let $P_{\mathsf{off}}(y|z \uparrow)$ be the probability that the AP is in OFF state at time $y, y \geq z$ given that a transition from OFF to ON period happens at time z. Because of memoryless property of the OFF period distribution,we do not need such notation for transition from ON to OFF period, since $P_{\mathsf{off}}(y|z \downarrow) = P_{\mathsf{off}}(y|z \in \mathsf{off})$. To further the analysis, we next find an expression for $P_{\mathsf{off}}(\cdot)$.

Lemma 3:

$$P_{\mathsf{off}}(t + x|t \in \mathsf{off}) = P(r_t \geq x) + \int_0^x f_d(z)P_{\mathsf{off}}(t + x|z \uparrow)dz, z \geq t \tag{3}$$

$$P_{\mathsf{off}}(t+x|t \uparrow) = \int_0^x f_c(w)P_{\mathsf{off}}(t+x|w \in \mathsf{off})dw, w \geq t. \tag{4}$$

Proof: The first expression is obtained by counting the two exclusive events, i) the residual time r_t of the present (at time t) OFF period exceeds x, and ii) the present OFF period expires at time z (i.e. OFF to ON transition happens at z), and taking the expectation of $P_{\mathsf{off}}(t + x|z \uparrow)$ with respect to $t \leq z \leq x$. The second expression follows similarly. ■

Corollary 4: For OFF period $\sim EXP(\lambda_d)$ and ON period $\sim EXP(\lambda_c)$, for $x \geq 0$,

$$P_{\mathsf{off}}(t + x|t \in \mathsf{off}) = \frac{\lambda_c}{\lambda_c + \lambda_d} + \frac{\lambda_d}{\lambda_c + \lambda_d}\exp(-(\lambda_d + \lambda_c)x).$$

Note that $P_{\mathsf{off}}(t + x|t \in \mathsf{off})$ does not depend on the starting time t as expected, because of the memoryless property of the exponential distribution.

Proof: We use the Laplace transforms of (3) and (4) to solve for $P_{\mathsf{off}}(t + x|t \in \mathsf{off})$. With OFF period $\sim EXP(\lambda_d)$ and ON period $\sim EXP(\lambda_c)$, $f_d(w) = \lambda_d \exp(-\lambda_d w)$ and $f_c(w) = \lambda_c \exp(-\lambda_c w)$. Denoting the Laplace transform of $f_d(w)$ as $f_d^*(s)$, $f_c(w)$ as $f_c^*(s)$, $P_{\mathsf{off}}(t+x|t \in \mathsf{off})$ as $\hat{P}_0^{of}(s)$ and $P_{\mathsf{off}}(t + x|t \uparrow)$ as $P_1^{of}(s)$, we have from (3) and (4),

$$\hat{P}_0^{of}(s) = \frac{1}{s + \lambda_d} + f_d^*(s)\hat{P}_1^{of}(s),$$
$$\hat{P}_1^{of}(s) = f_c^*(s)\hat{P}_0^{of}(s).$$

Note that $f_c^*(s) = \frac{\lambda_c}{s+\lambda_c}$ and $f_d^*(s) = \frac{\lambda_d}{s+\lambda_d}$. Hence

$$\hat{P}_0^{of}(s) = \frac{1}{s+\lambda_d}\frac{1}{1-f_d^*(s)f_c^*(s)}, \quad (5)$$

$$= \frac{1}{\lambda_d+\lambda_c}\left(\frac{\lambda_c}{s} + \frac{\lambda_d}{s+\lambda_d+\lambda_c}\right). \quad (6)$$

Taking the inverse transform, for $x \geq 0$

$$P_{\mathsf{off}}(t+x|t \in \mathsf{off}) = \frac{\lambda_c}{\lambda_c+\lambda_d} + \frac{\lambda_d}{\lambda_c+\lambda_d}\exp(-(\lambda_d+\lambda_c)x).$$

■

Next, we find the expected running cost $\mathbb{E}\{M(t, b(t))\}$ to compute the expected cost $\mathbb{E}\{c(t, b(t)\}$. Again appealing to the memoryless property of the exponential distribution, $\mathbb{E}\{M(t, b(t))\} = \mathbb{E}\{M(b)\}$ where we have shifted the starting time to 0. We will use recursions similar to (3) and (4) and Laplace transforms to find $\mathbb{E}\{M(t)\}$.

Let $\mathsf{M}_d(t) = \mathbb{E}\{M(t)\}$ be the average missed ON period time between times τ to $t+\tau$, when $\tau \in \mathsf{off}$. Moreover, let $\mathsf{M}_\uparrow(t) = \mathbb{E}\{M(\tau, \tau+t)\}$ be the average missed ON period time between times τ to $t+\tau$ given that the OFF to ON transition happens at time τ, and similarly let $\mathsf{M}_\downarrow(t) = \mathbb{E}\{M(\tau, \tau+t)\}$ given that the ON to OFF transition happens at time τ. The quantities $\mathsf{M}_\uparrow(t)$ and $\mathsf{M}_\downarrow(t)$ have no dependence on τ because of the Markov property of ON and OFF periods. So without loss of generality, we take $\tau = 0$. Note that because of memoryless property of OFF periods $\mathsf{M}_d(t) = \mathsf{M}_\downarrow(t)$.

Lemma 5:

$$\mathsf{M}_\uparrow(t) = t\int_t^\infty f_c(x)dx + \int_0^t f_c(x)(x + \mathsf{M}_\downarrow(t-x))dx, \quad (7)$$

and

$$\mathsf{M}_\downarrow(t) = \int_0^t f_d(x)\mathsf{M}_\uparrow(t-x)dx. \quad (8)$$

Proof: To derive (7), we have broken the expectation $\mathsf{M}_\uparrow(t)$ into two terms, where in the first we count the expected length of the ON period that starts at time $\tau = 0$ and continues beyond time t, and in the second, we consider the case when the ON period that starts at time $\tau = 0$ finishes at some time $x < t$ and count for the expected loss with ON to OFF transition happening at x. The second expression (8) follows similarly. ■

Theorem 6: For OFF period $\sim EXP(\lambda_d)$ and ON period $\sim EXP(\lambda_c)$, the expected loss $\mathsf{M}_d(t)$ is given by

$$\mathsf{M}_d(t) = \frac{\lambda_d}{\lambda_d+\lambda_c}\left(t - \frac{1-e^{-(\lambda_d+\lambda_c)t}}{\lambda_c+\lambda_d}\right). \quad (9)$$

Proof: We take the Laplace transforms of (7) and (8) to get

$$\mathsf{M}_\uparrow^*(s) = \frac{(f_c^*(0) - f_c^*(s))}{s^2} + f_c^*(s)\mathsf{M}_\downarrow^*(s),$$

and $\mathsf{M}_\downarrow^*(s) = f_d^*(s)\mathsf{M}_\uparrow^*(s)$. So we have

$$\mathsf{M}_d^*(s) = \mathsf{M}_\downarrow^*(s) = f_d^*(s)\frac{(f_c^*(0) - f_c^*(s))}{s^2(1-f_c^*(s)f_d^*(s))}.$$

Substituting for $f_c^*(s) = \frac{\lambda_c}{s+\lambda_c}$ and $f_d^*(s) = \frac{\lambda_d}{s+\lambda_d}$, and taking the inverse Laplace transform we obtain the result. ■

Remark 7: It is important to note that a similar derivation for $\mathsf{M}_d(t)$ has been attempted in [1], however, contains glaring miscalculations. For example, in [1] the ON period cost $\mathsf{M}_d(t)$ incurred in time t does not depend on t, and is always less than 1.

Finally, we have all the intermediate results to solve for the DP (2) when the OFF periods are exponentially distributed. With $b(t) = b$, the DP simplifies to

$$V(b) = \max_{b\geq 0}[\mathbb{E}\{c(b)\} + P_{\mathsf{off}}(b)V(b)]. \quad (10)$$

Theorem 8: The optimal sensing duration b satisfies the following equation

$$\frac{\mathrm{d}}{\mathrm{d}b}V = \frac{\mathrm{d}}{\mathrm{d}b}\left(\frac{c_s + \mathsf{M}_d(b)}{1-P_{\mathsf{off}}(b)}\right) = 0, \quad (11)$$

where $\mathsf{M}_\downarrow(b) = \mathsf{M}_d(b)$ can be found by substituting for $f_c(x)$ and $f_d(x)$ in (7) and (8) and $P_{\mathsf{off}}(b)$ can be found by substituting for $f_c(x)$ and $f_d(x)$ in (3) and (4).

Proof: Follows by rewriting (10), and taking the V terms common, and equating the derivative of V with respect to b to zero. ■

Corollary 9: For OFF period $\sim EXP(\lambda_d)$ and ON period $\sim EXP(\lambda_c)$, the optimal sensing duration b satisfies

$$e^{-(\lambda_c+\lambda_d)b}\left(1 + \frac{c_s}{\lambda_d}(\lambda_c+\lambda_d)^2 + b(\lambda_c+\lambda_d)\right) = 1. \quad (12)$$

Proof: From Corollary 4 and Theorem 6, substituting for $P_{\mathsf{off}}(b)$ and $\mathsf{M}_\uparrow(b)$ in (10), we get

$$V(b) = \frac{c_s + \frac{\lambda_d}{\lambda_d+\lambda_c}\left(b - \frac{1-e^{-(\lambda_d+\lambda_c)b}}{\lambda_c+\lambda_d}\right)}{\frac{\lambda_d}{\lambda_c+\lambda_d}\left(1-e^{-(\lambda_d+\lambda_c)b}\right)}. \quad (13)$$

Equating $\frac{\mathrm{d}}{\mathrm{d}b}V = 0$, we get (12). While this is a trascendental equation, the numerical solution is easy, see Figure 4. It is also easy to check that the second derivative of V is ≥ 0 and hence the above solution is indeed the global minimum. ■

In Figs. 4 and 5, we solve for (13) and plot the optimal cost V and the optimal sensing duration b. These are then compared against the values obtained by the blind learning algorithm proposed in Section VI.

Remark 10: We can find the optimal sensing duration b for other ON period distributions as long as the OFF period distribution is exponential via solving (11), after plugging in the ON period distribution.

Note that the use of dynamic programming framework helps in solving the problem using only the Laplace transform of the ON-period distribution. After considering the specific case of exponential distribution for the OFF periods, we generalize the results in next section, where general OFF period distribution is considered.

V. NON EXP-OFF PERIOD

Recall from Remark 1 that in the framework of Section III, we cannot solve for the optimal sensing durations when the OFF period distribution is not exponential. To circumvent this restriction, in this section we make an extra assumption following [3], where if any complete ON period is missed because of no sensing in that ON period, the mobile device is made aware of that and the system is reset. Thus, the problem (2) is now restricted to one OFF (X) and one ON (Y) period, and we want to choose the sensing durations so as to minimize the sum of the expected missed ON period time and the expected sensing cost.

Under this model, the DP in (2) becomes,

$$V(t) = \max_{b(t)\geq 0} [\mathbb{E}\{c(t, b(t))\} + P(r_t > b(t))V(t+b(t))], \quad (14)$$

where the cost function $c(t, b(t))$ is simplified and given by $c(t, b(t)) = c_s + \mathbb{E}\{(b(t) - r_t)\mathbf{1}_{b(t)\geq r_t, Y > b(t)-r_t}\} + \mathbb{E}\{Y\mathbf{1}_{(Y\leq b(t)-r_t)}\}$, where c_s is the fixed sensing cost. The lost ON period is written as two terms, either one complete missed ON period Y if $b(t)$ is larger than $r_t + Y$, otherwise, the missed ON period is $(b(t) - r_t)$.

Problem (14) is a generalization of problem considered in [8], where the length of the ON period is infinite (not a random variable) and the problem is to minimize the sum of the expected time lost in detecting ON period and the expected sensing cost. Note that in our setup, since the ON period expires in finite expected time, the solution of [8] does not apply.

We use a state space approach to derive results when the ON and OFF periods have a general distribution. The state space we consider is the set of non-negative real numbers. An action $b(t)$ is the duration of the next sleep period. We assume that $b(t)$ can take values only in a finite set (which is clearly true in practice). Thus, the set of t reachable (with positive probability) by any policy is countable and hence without loss of generality we assume that the state space is discrete. Then we have the following result.

Theorem 11: To solve (14), for any ON and OFF period distribution, the following statements hold.

1) There exists an optimal deterministic stationary policy.

2) Let $V^0 = 0$, $V^{k+1} = \mathcal{L}V^k$, where $\mathcal{L}V(t) = \min_{b(t)}[c(t, b(t)) + P(r_t > b(t))V(t + b(t))]$ and $c(t, b(t))$ is the per stage/running cost. Then V^k converges monotonically to the optimal value V^*.

3) V^* is the smallest nonnegative solution of $V^* = \mathcal{L}V^*$. A stationary policy that chooses at state (time) t an action that achieves the minimum of $\mathcal{L}V^*$ is optimal.

Proof: 1) follows from the [[11], Thm 7.3.6] that states that if the state space is discrete (finite or countable) and the action set for any state is finite, then there exists an optimal deterministic stationary policy. These conditions are satisfied in this case since the set of all actions (b's) is assumed to be finite, and the state space is countable. Similarly, 2) follows from the [[11], Thm 7.3.10] that states that if reward w for action a at state s, $w(s, a) \geq 0$, and state space is countable and action space is finite for each state, then if $V^0 = 0$, $V^{n+1} = \mathcal{L}V^n$ converges monotonically to V^* and 3) follows from [[11], Thm 7.3.3]. ∎

Theorem 11 shows that it is sufficient to consider deterministic policies without losing out on optimality, and randomized strategies are not needed. Moreover, part 2) and 3) tell us that the value iteration policy converges to the optimal solution for any ON and OFF period distributions.

We now consider the special case when the OFF period depends on the time at which it starts, but in the limit of very large time t, it loses that dependence.

Theorem 12: Assume that the residual OFF period r_t converges in distribution to r, and define $v(b) = \frac{c^\star(b)}{1-P(\mathsf{r}>b)}$. Then

1) $\lim_{t\to\infty} V^*(t) = \min_b v(b)$.

2) Assume that there is a unique b that achieves the minimum of $v(b)$ and denote it by $b^\star$. Then there is some stationary optimal policy $b(t)$ such that for all t large enough, $b(t) = b^\star$.

Proof: Let $V^0 = 0$, and assume that $\bar{V}^k = \lim_{t\to\infty} V^k(t)$ exists for some k. Then from definitions used in Theorem 11, we have

$$\begin{aligned}\bar{V}^{k+1} &= \lim_{t\to\infty} \mathcal{L}V^k(t),\\ &= \lim_{t\to\infty} \min_{b(t)}[c(t, b(t)) + P(r_t > b(t))V^k(t + b)],\\ &= \min_b[c^\star(b) + P(\mathsf{r} > b)\bar{V}^k],\end{aligned}$$

where the last equality follows since r_t converges in distribution to r, and from the bounded convergence theorem

$$\lim_{t\to\infty} c(t, b(t)) \to c^\star(b).$$

Essentially, since r_t converges in distribution to r, the per-stage/running cost $c(t, b(t))$ becomes independent of t as $t \to \infty$ (similar to the case when OFF periods are exponentially distributed). Hence by convergence of V^k to V^* by Theorem 11, the limit $\bar{V} = \lim_{t\to\infty} V^*(t)$ exists. Thus, there exists a constant deterministic policy as $t \to \infty$ which we denote by $b^\star$. This gives $\bar{V} = c^\star(b) + P(\mathsf{r} > b)\bar{V}$ which on rearranging gives (i). (ii) can be obtained by noting that $b^\star$ performs better than any other policy so the optimal solution $b(t)$ must tend to $b^\star$ as $t \to \infty$ ∎

Therefore, if the residual OFF period distribution converges in time, then the optimal sensing duration converges to a constant after sufficiently long time.

Remark 13: The condition in Theorem 12 is trivially true for exponentially distributed OFF period. A more non-trivial example is when the OFF period has hyper-exponential distribution, for which the residual OFF period r_t converges in distribution to some r.

Remark 14: Theorem 11 and Theorem 12 are similar to Propositions III.2 and III.3 in [8].

A. Example

Next, we consider an example where both the OFF and ON periods are uniformly distributed between $[0, L_f]$ and $[0, L_o]$, respectively. To use Theorem 11, with sensing duration $b(t)$, we write down the sensing cost $c(t, b(t))$ and the probability $P(r_t > b(t))$ that at the next sensing epoch we again encounter an OFF period.

For OFF period distributed uniformly between $[0, L_f]$, the residual OFF period distribution

$$F_{r_t}(x) = P(r_t > x) = P(X > x | X > t) = \frac{L_f - x}{L_f - t}, \quad (15)$$

for $t \leq x \leq L_f$. Thus, $P(r_t > b(t)) = \frac{L_f - b(t)}{L_f - t}$, and to compute cost $c(t, b(t))$, we calculate

$$\mathbb{E}\{(b(t) - r_t)\mathbf{1}_{b(t) \geq r_t, Y > b(t) - r_t}\} = \frac{\frac{L_o^2}{6} + (1 - b(t))\frac{L_o}{2}}{L_f - t},$$

and

$$\mathbb{E}\{Y \mathbf{1}_{(Y \leq b(t) - r_t)}\} = \frac{b(t)^2 - b(t)(L_f - t) + \frac{L_f^2 + L_f t + t^2}{3}}{2L_o},$$

where the total cost $c(t, b(t)) = c_s + \mathbb{E}\{(b(t) - r_t)\mathbf{1}_{b(t) \geq r_t, Y > b(t) - r_t}\} + \mathbb{E}\{Y \mathbf{1}_{(Y \leq b(t) - r_t)}\}$.

Hence to solve for the optimal sensing durations $b(t)$ via the value iteration method, we start with $V = 0$, and write $V^{k+1} = \mathcal{L}V^k$, where

$$\mathcal{L}V(t) = \min_{b(t)} [c(t, b(t)) + P(r_t > b)V(t + b(t))],$$

where we substitute for $c(t, b(t))$ from above. From Theorem 11, these iterations converge to the optimal policy.

In this section, we considered the general ON-OFF period distributions, and derived some important structural results. Our results say that the optimal policy is deterministic, which is useful for practical purposes. Moreover, we also show that an optimal policy can be found via value iteration which is guaranteed to converge to the optimal solution. In the next section, we consider a more general regime, where we present a learning framework, in which the algorithm has no prior information about the ON-OFF period distributions, and the algorithms objective is to minimize the loss between itself and the optimal algorithm which has the knowledge of the two distributions.

VI. Learning Framework

In Sections IV and V, we have derived the optimal sensing duration assuming the knowledge of the distribution of the OFF and ON periods. In practice, learning these distributions is a problem in its own right. To obviate the need for exactly learning the distribution (might take a long training time), in this section, under the general model of Section III, we present a continuous-armed bandit problem type formulation [9], [10], where the algorithm learns the best sensing duration without explicitly knowing the underlying OFF and ON period distribution. For ease of exposition, we will assume that the OFF and ON periods are exponentially distributed with unknown parameters. The analysis carries over to all distributions for which the cost V has continuous second derivatives.

An online learning algorithm chooses one possible sensing duration b for each session (defined earlier), and receives a reward that counts for the useful ON period and the cost incurred (lost ON period time and sensing cost). Depending on past choices of b, the algorithm modulates its choice of b in future sessions in pursuit of larger rewards.

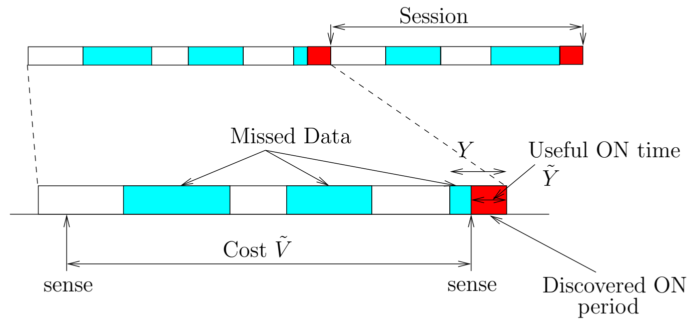

Fig. 3. Illustration of sessions for learning algorithm

The reward in session i is $U_i = \tilde{Y} - \tilde{V}$, where as shown in Fig. 3, $\tilde{Y}$ is the length of the useful ON period (or discovered ON period), and $\tilde{V}$ is the random variable whose expectation is cost (13), that counts the sensing cost and missed ON periods in each session. Note that there could be multiple sensing epochs in each session, and the sensing cost of each session is c_s times the number of sensing epochs at which an OFF period is sensed in that session. Let OFF period $\sim EXP(\lambda_d)$ and ON period $\sim EXP(\lambda_c)$, with unknown parameters λ_d and λ_c, where we assume that λ_d and λ_c are such that the optimal $b^\star \in [0, b_{max}]$ from Corollary 9.

The online algorithm's objective is to minimize the expected regret,

$$\min_{b(i), i=1,\ldots,T} \mathbb{E}\{R(T)\},$$

by choosing action $b(i)$ is session i, and

$$R(T) = \sum_{i=1}^{T} U_i^\star - \sum_{i=1}^{T} U_i, \quad (16)$$

where $U_i^\star$ is the optimal reward knowing λ_d and λ_c, i.e. playing optimal b from Corollary 9 in each session, and T is the time horizon.

The learning algorithm called the **OnlineLearning** [10] to find the sensing duration b to minimize the expected regret is given at the top of the page.

Lemma 15: (Theorem 1 [10]) If the expected reward given a strategy has continuous second derivatives, and finite number of maximas, then the expected regret obtained by the **OnlineLearning** algorithm is bounded as follows,

$$\mathbb{E}\{R(T)\} \leq \mathcal{O}\sqrt{T \log T},$$

for $n = \left(\frac{T}{\log T}\right)^{1/4}$.

	OnlineLearning
1	Choose n
2	Divide $[0, b_{max}]$ into n intervals $I_k = b_{max}[\frac{k-1}{n}, \frac{k}{n}]$, $0 \le k \le n$
3	For each I_k, choose a point (sensing duration b) uniformly at random
4	For $i = 1 : T$
5	Choose that interval I_k that maximizes $\hat{U}_k + \sqrt{\frac{2 \ln i}{t_k}}$, where $\hat{U}_k$ is the average (empirical) reward obtained from points in interval I_k so far, and t_k is the number of times interval I_k has been chosen till session i and i is the overall number of sessions so far
6	Choose a point uniformly at random from the chosen interval I_k.

Thus, the average regret $\left(\frac{\mathbb{E}\{R(T)\}}{T}\right)$ converges to zero with the **OnlineLearning** algorithm even without knowing the underlying distributions.

Theorem 16: Using the **OnlineLearning** algorithm, the normalized regret

$$\frac{\min_{b(i), i=1,\dots,T} \mathbb{E}\{R(T)\}}{T}$$

goes to zero with increasing number of sessions for finding the optimal sensing duration without the knowledge of OFF and ON period parameters λ_d and λ_c.

Proof: Note that the expectation V (13) of the cost function $\tilde{V}$ has continuous second derivatives and finite number of maximas for a fixed strategy b, and $\mathbb{E}\{\tilde{Y}\}$ does not depend on b because of memoryless property of exponential distribution. Thus, the expected reward $\mathbb{E}\{U_i\}$ given b has continuous second derivatives and finite number of maximas, and the result follows from Lemma 15. ∎

VII. Simulations

In Fig. 4, we demonstrate the performance of the **OnlineLearning** via simulation for EXP ON-OFF periods and compare it with that of the optimal solution (13) that knew the parameters of the two distributions. We use λ_d and λ_c such that the expected OFF period length and ON period length is 3 and 2, respectively, and plot the optimal sensing duration and sensing duration discovered by **OnlineLearning** algorithm as function of the sensing cost c_s. We see that the **OnlineLearning** algorithm closely tracks the theoretical optimum computed by Corollary 9. Furthermore, the average cost incurred while connecting to the AP, closely matches for the two algorithms, as shown in Fig. 5, which is consistent with our theoretical result on asymptotic zero regret.

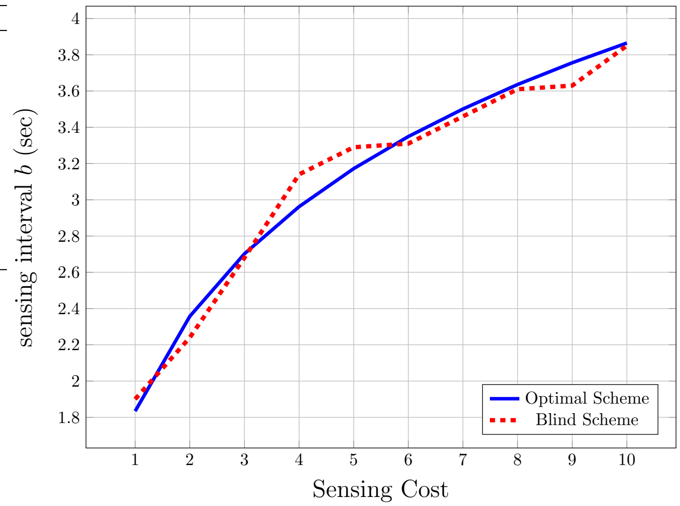

Fig. 4. Optimal sensing interval vs sensing cost

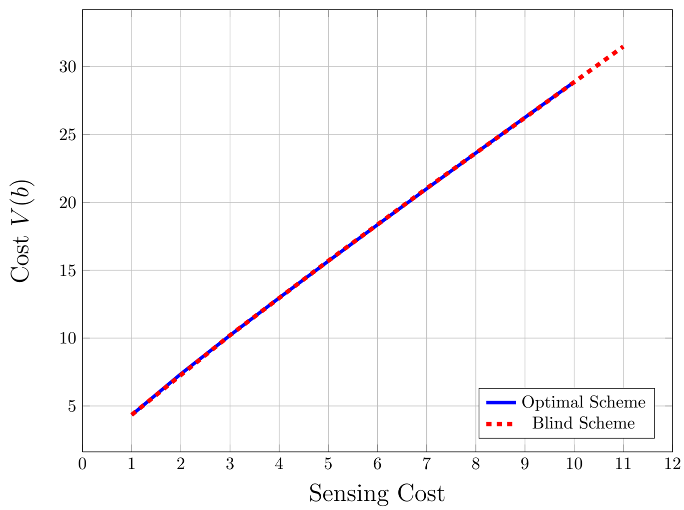

Fig. 5. Average cost vs sensing cost

VIII. Conclusions

In this paper, we found optimal sensing durations for mobile devices when both the lost data transmission opportunity and sensing cost are at the premium. First we concentrated on the case when the OFF period durations are exponentially distributed, which has been confirmed to be accurate using experimental data in prior work. In this setting, we showed that the sensing durations are fixed and deterministic, and can be found via solving a transcendental equation given the rate of the exponential distribution for the OFF period and the Laplace transform of the ON period. We also looked at the more general regime when the OFF period duration is not necessarily exponentially distributed, where we again showed that the optimal sensing duration is deteministic, and can be found via value iteration of the suitable cost function. We also derived some limiting results for this case when the residual OFF period converges in distribution. Similar analysis has been attempted in past, but either has some inconsistencies [1] or uses a simplified cost function to facilitate the analysis [3]. Finally, to completely obviate the need for learning the OFF and ON period distributions, we considered the learning framework where the explicit information about the OFF and ON period is not required. The learning algorithm is based on continuous bandit learning algorithms, where the algorithm tries to minimize the regret, and is shown to achieve zero regret asymptotically. Thus, even without the knowledge of distribution, the algorithm does as well as the optimal algorithm that knows the distributions exactly.

REFERENCES

[1] K.-H. Kim, A. W. Min, D. Gupta, P. Mohapatra, and J. P. Singh, "Improving energy efficiency of Wi-Fi sensing on smartphones," in *INFOCOM, 2011 Proceedings IEEE*. IEEE, 2011, pp. 2930–2938.

[2] N. LaMarca and J. Hightower, "CRAWDAD trace intel/placelab/ placelab/downtown (v. 2004-12-17)," 2004.

[3] J. Jeong, Y. Yi, J. Cho, D. Eun, and S. Chong, "Wi-Fi sensing: Should mobiles sleep longer as they age?" in *INFOCOM, 2013 Proceedings IEEE*. IEEE, 2013, pp. 2328–2336.

[4] W. Wang, M. Motani, and V. Srinivasan, "Opportunistic energy-efficient contact probing in delay-tolerant applications," *IEEE/ACM Transactions on Networking (TON)*, vol. 17, no. 5, pp. 1592–1605, 2009.

[5] R. Zhou, Y. Xiong, G. Xing, L. Sun, and J. Ma, "ZiFi: wireless lan discovery via zigbee interference signatures," in *Proceedings of the sixteenth annual international conference on Mobile computing and networking*. ACM, 2010, pp. 49–60.

[6] H. Wu, K. Tan, J. Liu, and Y. Zhang, "Footprint: cellular assisted Wi-Fi ap discovery on mobile phones for energy saving," in *Proceedings of the 4th ACM international workshop on Experimental evaluation and characterization*. ACM, 2009, pp. 67–76.

[7] Y. Chen, Q. Zhao, and A. Swami, "Distributed spectrum sensing and access in cognitive radio networks with energy constraint," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 783–797, Feb 2009.

[8] A. P. Azad, S. Alouf, E. Altman, V. Borkar, and G. S. Paschos, "Optimal control of sleep periods for wireless terminals," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1605–1617, 2011.

[9] R. D. Kleinberg, "Nearly tight bounds for the continuum-armed bandit problem," in *Advances in Neural Information Processing Systems*, 2004, pp. 697–704.

[10] P. Auer, R. Ortner, and C. Szepesvári, "Improved rates for the stochastic continuum-armed bandit problem," in *Learning Theory*. Springer, 2007, pp. 454–468.

[11] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2009, vol. 414.