Closed and Open Loop Optimal Control of Buffer and Energy of a Wireless Device

V. S. Borkar
School of Technology and Computer Science
TIFR, Mumbai, India.
borkar@tifr.res.in

A. A. Kherani B. J. Prabhu
INRIA Sophia Antipolis
06902 Sophia Antipolis, France.
{alam,bprabhu}@sophia.inria.fr

***Abstract*—We study a decision problem faced by an energy limited wireless device that operates in discrete time. There is some external arrival to the device's transmit buffer. The possible decisions are: a) to serve some of the buffer content; b) to order a new battery after serving the maximum possible amount that it can; or c) to remain idle so that the battery charge can increase owing to diffusion process (which is possible in some commercially available batteries). We look at open and closed-loop controls of the system. The closed-loop control problem is addressed using the framework of Markov Decision Processes. We consider both finite and infinite horizon discounted costs as well as average cost minimization problems. *Without using any second order characteristics*, we obtain results that include i) optimality of bang-bang control, ii) the optimality of threshold based policies, iii) parametric monotonicity of the threshold, and iv) uniqueness of the threshold. For the open-loop control setting we use recent advances in application of multimodular functions to establish optimality of bracket sequence based control.**

I. Introduction

Wireless devices are constrained in their operational lifetime by finite energy batteries. Therefore, energy efficient design of protocols at different layers of the protocol stack for wireless networks has recently received significant attention, see, for example, [1]-[4]. Although the primary objective of a terminal is to transmit and receive data with minimal delay, this must be done with the added constraint of minimizing the transmission costs and increasing the operational lifetime of the terminal. In [5], the authors studied delay optimal packet scheduling policies subject to average transmit power constraint over a wireless channel with independent fading. In [6], the authors extended this model to include Markovian fading. Although in the above mentioned articles an average transmit power constraint was imposed, an interesting feature of the battery was ignored. In [7], it was observed that a battery when left idle can regain some of its lost charge. This phenomenon, known as relaxation phenomenon, allows a battery operated in pulsed (or intermittent) discharge mode to deliver more energy than the same battery operated in continuous discharge mode. This enables a user to send more packets and increase the operational lifetime of the terminal by leaving the battery idle in between packet transmissions thus providing an incentive to remain idle even though the transmit buffer may not be empty. However, this would add to the delay of the packets queued up in the buffer. This trade-off between energy and delay leads to a decision making problem formulation where the user has to decide whether to serve packets or leave the terminal idle in order to minimize certain costs.

In this paper, we consider a discrete time system in which a user with a finite energy battery terminal has to decide whether to serve packets or to leave the system idle in each time slot. Further, the user can decide to replace the battery with a new one at an additional cost. We note that there are two variables (i.e., energy level of the battery and the length of the transmit buffer) based on which a decision is to be made. We formulate the problem as a Markov decision process. We then derive the structural properties based on the directional derivative of the value function. We first consider a finite horizon problem and provide the structure of the optimal policy. We then extend this to the infinite horizon discounted cost problem, and finally consider the infinite horizon average cost minimization. We then consider the problem of making an optimal decision when the knowledge of the remaining energy and buffer occupancy are not known.

The outline of the paper is as follows. In Section II, we formulate the problem of closed loop control. Sections III IV-B and IV-C deal with finite horizon discounted cost, infinite horizon discounted cost and infinite horizon average cost, respectively. Section VI deals with open loop control. For the proofs, and the definitions used in Section VI on open loop control, the reader is referred to the detailed research report [8].

II. Closed Loop Control

We first consider the optimal control problem where, at the beginning of each time slot (i.e., decision epoch), the device is aware of the current buffer occupancy and the energy level of the battery, and hence can take an action based on these two parameters. However, it does not have any knowledge of the amount of data that will arrive in the transmit buffer during the current time slot. In this section, we shall formalize the problem statement mentioned in the Introduction.

A. Problem Formulation

Let x_n and p_n denote the buffer length and the remaining energy level, respectively, at the beginning of the n^{th} time slot. We assume x_n is infinitely divisible and $x_n \in [0, \infty), \forall n$, i.e., the buffer content is fluid and there is infinite buffer space. The remaining energy level, p_n, is assumed to be bounded above by M, i.e., $p_n \in [0, M], \forall n$. A linear relationship is assumed between the amount of energy and the amount of fluid that can be transmitted using this energy. In other words, a unit of energy is required to serve each unit of fluid. The state space, $\mathcal{C}$, is given by $\mathcal{C} = \{(x,p) : x \in [0,\infty), p \in [0,M]\}$. The cost function associated with the state (x,p) is denoted by $h(x) + g(p)$, where $h(x)$ is an increasing function of x and $g(p)$ is a decreasing function of p. We could also use a composite cost function $c(x,p)$ instead of $h(x) + g(p)$, and all the results in the paper would continue to hold. However, for simplicity, we use the above mentioned form. Let w_n denote the amount of fluid which arrives during the n^{th} time slot. The random sequence $\{w_n\}, n \geq 0$, is assumed to be composed of independent and identically distributed random variables, each with distribution function $\mu(\cdot)$.

We assume that when the battery is left idle in a slot, the residual battery energy (or, charge) increases from p to some amount $p + B(p) \geq p$. In the rest of the article, we will drop the dependence on p of $B(p)$ and use B to denote the function. We note that the case $B = 0$ corresponds to the other practical scenario where the battery does not gain its charge when left idle. The following development easily extends to the case where $B(p) < 0$. In state (x,p), the user can take one of the following actions:

1) remain idle,
2) serve some amount $u \in [0, x \wedge p]$ without reordering a new battery, or
3) serve $x \wedge p$ and reorder a new battery with residual energy level M.

Here the symbol $\wedge$ denotes the minimum operator. We denote the action space by $\mathcal{A}$, where $\mathcal{A} = \{1,2,3\}$. A cost of $r(p)$, where $r(\cdot)$ is a non-decreasing function of p, is incurred each time a battery is reordered in state (x,p). A policy π defines an action for each $(x,p) \in \mathcal{C}$. Let β denote the discount factor. We study optimal policies which minimize one of the following three cost criteria.

- Finite horizon discounted cost
$$\sum_{k=0}^{N} \beta^k (h(x_k) + g(p_k)),$$
for some $N > 0$.
- Infinite horizon discounted cost
$$\sum_{k=0}^{\infty} \beta^k (h(x_k) + g(p_k)).$$
- Infinite horizon average cost
$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N} (h(x_k) + g(p_k)).$$

III. The Finite Horizon Discounted Cost Case

Let $V_k(x,p)$ denote the cost when the system state is (x,p) and there are k more decision epochs before reaching the horizon. Since the decision epoch can be determined from V_k, we use x, p and w instead of x_k, p_k and w_k, respectively. For the finite horizon discounted cost case, the dynamic programming equation is

$$\begin{aligned} V_{k+1}(x,p) &= h(x) + g(p) \\ &+ \beta \min \Big(E_w V_k((x-p)^+ + w, M) + r((p-x)^+), \\ &\qquad E_w V_k(x+w, (p+B) \wedge M), \\ &\qquad \min_{0 \leq u \leq x \wedge p} E_w V_k(x+w-u, p-u) \Big). \quad (1) \end{aligned}$$

The expectation operator, E_w, is the expectation over the random variable w. The first term for the minimum operator corresponds to the decision of serving the maximum possible fluid and reordering a new battery. The second term corresponds to the decision of leaving the battery idle, and the third term correponds to the decision of serving some amount of fluid. An important property satisfied by the above formulation is that

$$\frac{\partial}{\partial p} r((p-x)^+) + \frac{\partial}{\partial x} r((p-x)^+) = 0.$$

In the rest of this section, we shall use a fixed cost c_3 for battery reordering instead of $r(\cdot)$.

In the rest of this section, we assume that $\beta = 1$ with a note that all the results of this section are valid when we consider a discount factor $\beta < 1$. Let N denote the finite horizon. We first provide a simple condition under which *bang-bang* control is optimal, i.e., the decision is to either serve the maximum possible quantity or to remain idle. We then provide structural result of the optimal policy, like parametric monotonicity of the threshold policy. We would like to point out here that we obtain the results of this section *without using second order characteristics of the value functions* such as convexity or decreasing-differences.

A. Optimality of Bang-Bang Control

Let $\nabla V(x,p) = \frac{\partial}{\partial y} V(y,p)|_{y=x} + \frac{\partial}{\partial q} V(x,q)|_{q=p}$ denote the directional derivate of $V(y,q)$ along the vector $(1,1)$ at (x,p). We now give a condition which is used to derive the structure of the optimal policy.

Assumption 1: The function $h'(x) + g'(p)$ has the same sign, say $\mathcal{S} \in \{+,-\}$, for all values of x and p.

Remark 1: The cost function $g(p)$ is a continuous non-increasing function defined on the closed interval $[0,M]$. The derivative of $g(p)$ is, therefore, negative and bounded from below. If $h(x)$ is a polynomial in x such that the coefficient of x is greater than $\sup_p |g'(p)|$ then $\mathcal{S}$ will be positive for all x and p. Similary, if the derivative of $h(x)$ is upper bounded by $\inf_p |g'(p)|$ then $\mathcal{S}$ will be negative for all x and p. For example, for $c_1 > 0$, the function $h(x) = c_1(1 - \exp(-x))x$ is convex and increasing with $h'(x)$ upper bounded by c_1.
This condition is quite general and is satisfied by many natural candidates for the cost functions $h(x)$ and $g(p)$. In particular, this condition is also satisfied by the linear functions $h(x) = c_1 x$ and $g(p) = c_2 p$.

Let

$$u^* := \operatorname*{argmin}_{0 \leq u \leq x \wedge p} E_w V_k(x+w-u, p-u)$$

denote the optimal amount of fluid to be served in a slot in which the decision is to serve some fluid. Under Assumption 1, we have the following two lemmas.

Lemma 1: For all values of x, p, w, and k, u^* is given by

$$u^* = \begin{cases} 0 & \text{if } \nabla V_k(x,p) < 0, \ \forall (x,p), \\ x \wedge p & \text{if } \nabla V_k(x,p) > 0, \ \forall (x,p). \end{cases}$$

The above lemma helps us to quantify the optimal amount of fluid to be served in a slot in which the decision is to serve some fluid. Depending on the sign of $\nabla V_k(x,p)$, the optimal amount of fluid to be served is either zero or the maximum possible amount. We now characterize the behaviour of the sign of $\nabla V_k(x,p)$.

Lemma 2: Under Assumption 1, the directional derivative $\nabla V_k(x,p)$ has the same sign $\mathcal{S}$ for all values of $k < N, x$, and p.

Corollary 1: For all values of x, p, w, and $k < N$, the optimal amount of fluid to be served, u^*, is

$$u^* = \begin{cases} 0 & \text{if } \mathcal{S} \text{ is negative,} \\ x \wedge p & \text{if } \mathcal{S} \text{ is positive.} \end{cases}$$

Using induction, one can prove the following lemma.

Lemma 3: $V_k(x,p)$ is decreasing in p for a fixed x and increasing in x for a fixed p.

Using the previous three lemmas, we can now obtain certain characteristics of the optimal policy for two different values of $\mathcal{S}$.

Theorem 1: Let $r(\cdot)$ be a constant, equal to c_3.

1) If $\mathcal{S}$ is negative then the optimal policy is to either
   a) serve maximum possible amount ($x \wedge p$) and **reorder**, or
   b) remain idle.
2) If $\mathcal{S}$ is positive then the optimal policy is to either
   a) serve maximum possible amount ($x \wedge p$) and **reorder**, or
   b) serve maximum possible amount ($x \wedge p$) and **do not reorder**, or
   c) remain idle.

For either value of $\mathcal{S}$, the optimal policy is *bang-bang* type, i.e., either no fluid is served or maximum possible amount of fluid is served.

Theorem 2: If $\mathcal{S}$ is positive and p is equal to M then the optimal policy is to first serve $x \wedge M$ and then decide to reorder or not.

The battery charge p can not increase beyond M. When $\mathcal{S}$ is positive and p is equal to M, leaving the terminal idle will add to the cost of increasing the delay of packets. Therefore, it is optimal not to remain idle.

For the special case where B is equal to zero and $\mathcal{S}$ is positive, we can strengthen Theorem 1 in the following way.

Theorem 3: Let $r(\cdot)$ be a constant, and let B be zero. When $\mathcal{S}$ is positive, the optimal policy is to either serve maximum possible amount ($x \wedge p$) and **reorder**, or serve maximum possible amount ($x \wedge p$) and **do not reorder**.

Remark When B is equal to zero and $\mathcal{S}$ is positive, it is optimal at all decision epochs to serve $x \wedge p$. We are, therefore, left with the decision to **reorder** or **not to reorder**. However, when B is equal to zero and $\mathcal{S}$ is negative, we can not eliminate the decision to remain idle from the action space. This point will become clear later when we consider infinite horizon discounted cost problem.

The above series of results have systematically reduced the number of choices to be made. We have also noted the optimality of bang-bang control policy, i.e, it is optimal to either serve the maximum possible amount or to serve nothing. Next we provide some structural results for the optimal policy. We study separately the case when $\mathcal{S}$ is negative and the case when $\mathcal{S}$ is positive.

IV. When $\mathcal{S}$ is negative

In this section we consider the case when $\mathcal{S}$ is negative and obtain structural results for finite and infinite horizon discounted cost, and then use *vanishing discount* approach to study the infinite horizon average cost case.

A. Structure of the Optimal Policy

We show that for each given value of p and k, there is a value $x_k^*(p)$ such that if $x < x_k^*(p)$ then the optimal action is to remain idle. That is, for each value of p and k there exists a threshold $x_k^*(p)$ such that if the amount of fluid in the buffer is less than the threshold then the optimal action is to remain idle. We also show that $x_k^*(p)$ is increasing function of p. First, we have the following lemma which can be proved using induction.

Lemma 4: The partial derivative $\frac{\partial}{\partial p} V_k(x,p)$ (resp. $\frac{\partial}{\partial x} V_k(x,p)$) is bounded above (resp. below) by $\max_p \frac{d}{dp} g(p)$ (resp. $\min_x \frac{d}{dx} h(x)$) for each $k \leq N$.

Remark If $\frac{d}{dp} g(p) < -1$ then $\frac{d}{dp} V_k(x,p) < -1$ for all values of $k \leq N$.

Since we are considering the case when $\mathcal{S}$ is negative, from Theorem 1, we only have two actions to choose from. Therefore, the dynamic programming equation simplfies to

$$\begin{aligned} V_k(x,p) &= h(x) + g(p) \\ &\quad + \min \big(E_w V_{k-1}((x-p)^+ + w, M) + c_3, \\ &\qquad E_w V_{k-1}(x+w, (p+B) \wedge M) \big). \quad (2) \end{aligned}$$

Lemma 5: At each decision epoch k, there is an $x_k^*(p)$ such that it is optimal to remain idle when $x < x_k^*(p)$ when the battery level is p.

Theorem 4: If $\mathcal{S} = -$, i.e., the directional derivative $\frac{d}{dx} h(x) + \frac{d}{dp} g(p)$ is less than 0, then $x_k^*(p)$ is an increasing function of p.

Remark In order to obtain the parametric monotonicity we have not used convexity or decreasing differences property of the value function (which, in fact, are not present in our case). Similarly we get

Corollary 2: If $\mathcal{X}(p)$ denotes the set of queue lengths x such that optimal decision is to remain idle when battery level is p, then $\mathcal{X}(p)$ is increasing in p in the sense that $\mathcal{X}(p) \subset \mathcal{X}(p+\delta)$, $\delta > 0$.

B. Infinite Horizon Discounted Cost

Now we consider the case of $N = \infty$, i.e., the infinite horizon problem. It is clear that all the properties obtained for the finite horizon problem are valid for this case also. For the finite horizon case, it was necessary to study the value function for all possible values of its argument, i.e., x and p. However, since we intend to study the average cost problem via the infinite horizon discounted cost case by using the *vanishing discount* approach, we will see that, for the case where $\mathcal{S}$ is negative, it is enough to study the value function $V(x,p)$ only at p equal to M. In this section, we study the structure of the infinite horizon discounted problem value function $V(x,M)$ assuming $\mathcal{S}$ is negative. We know that $V(\cdot,\cdot)$ satisfies the dynamic programming equation

$$\begin{aligned} V(x,p) &= h(x) + g(p) \\ &\quad + \beta \min \Big(E_w V((x-p)^+ + w, M) + c_3, \\ &\qquad E_w V(x+w, (p+B)\wedge M), \\ &\qquad \min_{0\le u\le x\wedge p} E_w(x-u+w, p-u) \Big). \end{aligned} \tag{3}$$

Assuming that $\mathcal{S}$ is negative, we can use value iteration for the above problem to show that

$$\begin{aligned} V(x,p) &= h(x) + g(p) \\ &\quad + \beta \min \Big(E_w V((x-p)^+ + w, M) + c_3, \\ &\qquad E_w V(x+w, (p+B)\wedge M) \Big). \end{aligned} \tag{4}$$

Assume now that $h(x)$ is a concave differentiable increasing function so that $\sup_x \frac{d}{dx}h(x) = \frac{d}{dx}h(x)|_{x=0} \le \frac{(1-\beta)c_3}{M}$. Now we consider the value-iteration for $p = M$, i.e., assume $V_0(x,p) = h(x) + g(p)$, and apply the above minimization iteratively, generating a family of value functions $V_k(x,p),\ k \ge 0$ so that

$$\begin{aligned} V_{k+1}(x,p) &= h(x) + g(p) \\ &\quad + \beta \min \Big(E_w V_k((x-p)^+ + w, M) + c_3, \\ &\qquad E_w V(_k x+w, (p+B)\wedge M) \Big). \end{aligned} \tag{5}$$

Since p is equal to M,

$$\begin{aligned} V_{k+1}(x,M) &= h(x) + g(M) \\ &\quad + \beta \min \Big(E_w V_k((x-M)^+ + w, M) + c_3, \\ &\qquad E_w V_k(x+w, M) \Big). \end{aligned} \tag{6}$$

We observe from the above expression that *once the battery energy reaches M, it stays there*. This simplifies the problem significantly as now, with initial battery level at M, the value function can be viewed as a function of x, the buffer occupancy, only. We now prove that

Theorem 5: The partial derivative of the value function with respect to x at x equal to zero, $\frac{\partial}{\partial x}V_k(x,M)|_{x=0}$, is less than or equal to $\frac{(1-\beta^{k+1})c_3}{M}$, and $V_k(x,M)$ is concave in x for all $k \ge 0$.

Remark: The above result suggests that if $\sup_x \frac{d}{dx}h(x) = \frac{d}{dx}h(x)|_{x=0} \le \frac{(1-\beta)c_3}{M}$ then it is optimal to remain idle *forever* whenever the battery is fully charged. This result, though seemingly counter-intuitive, can be explained by the fact that we are considering discounted cost which gives weight to near future cost only. To be able to give more consideration to distant future, we require β very close to unity, in which case the condition of the theorem ($\sup_x \frac{dh(x)}{dx} = \frac{dh(x)}{dx}|_{x=0} \le \frac{(1-\beta)c_3}{M}$) does not hold. This point will be clearer when we consider the average cost optimization problem where one gives all weight to the distant future costs.

Remark: If $h(x)$ is not concave then we can only say that $\frac{\partial}{\partial x}V_{k+1}(x,M)|_{x=0} \le \frac{(1-\beta)c_3}{M}$.

If we consider a linear form for the buffer cost, i.e., $h(x) = c_1 x$ then we have that

Theorem 6: If $h(x) = c_1 x$ with $c_1 \le \frac{(1-\beta)c_3}{M}$ then

$$\frac{\partial}{\partial x}V_k(x,M)|_{x=0} = \frac{1-\beta^{k+1}}{1-\beta}c_1,$$

so that $V_k(x,M)$ is linear in x for all $k \ge 0$.

Corollary 3: If $h(x) = c_1 x$ with $c_1 \le \frac{(1-\beta)c_3}{M}$ then $V(x,M) = \frac{c_1 x + g(M)}{1-\beta}$.

Theorem 7: If $\mathcal{S}$ is negative and if $h(x)$ is concave function such that and $\inf_x h'(x) = (1-\beta)(\frac{c_3}{M} + L)$ for $L > 0$ then there exists an N such that $V_N(x,M) > V_N(0,M) + c_3$ for some $x < M$.

Note that once the battery attains its maximum capacity, i.e., M, then if $\mathcal{S}$ is negative, the battery always remains fully charged. Hence, if the intial battery charge is M then we can consider $V_n(x,p)$ as a function of x alone (when considering the infinite horizon problem). For this case we have the result of Theorem 8 which requires the following definition.

Definition We say that a function $f(x)$ is M-convex if

$$f'(x) \le f'(x+M),\ \forall x.$$

Theorem 8: If $\mathcal{S}$ is negative, and $h(x)$ is M-convex and increasing then $V_n(x,M)$ is continuous and M-convex, i.e., $\frac{d}{dx}V_n(x+M) \ge \frac{d}{dx}V_n(x), \quad x \ge 0$.

Remark The above result also implies that $V_n(x,M)$ is neither convex nor concave in general.

Now we assume, without loss of generality, that $\inf_x h'(x) = 1$ (this can always be done by appropriately scaling $g(p)$ and c_3). We now use result from Lemma 4 to provide structure of the optimal policy.

Theorem 9: If $\mathcal{S}$ is negative, and $h(x)$ is M-convex and increasing with $h'(0) = 1$ and $c_3 < M$ then there is a unique threshold T such that, for $p = M$, if $x \le T$ then it is optimal to remain idle for the infinite horizon problem else it is optimal to serve $x \wedge M$ and reorder the battery.

Theorem 7 can now be applied to the case where $h(x) = c_1 x$ and then Theorem 9 can be used to obtain more structural results when $h(x) = c_1 x$. We thus have the following structure when $\mathcal{S}$ is negative and $h(x) = c_1 x$ (when starting from $p = M$):

1) if $c_1 \le (1-\beta)\frac{c_3}{M}$ then it is optimal to always remain idle when using discount factor of β
2) else there is a $T < c_3$ such that it is optimal to remain idle for $x < T$ and reorder battery for $x > T$.

The first result is obtained from Theorem 5 and the second part is obtained as follows: since $c_1 > (1-\beta)\frac{c_3}{M}$, starting from

$p = M$, in the value iteration we will ultimately get a stage at which $V(x, M) > V(0, M) + c_3$ for some $x < M$. Now, since the structure derived for $h(x)$ convex is valid here, Theorem 9 can be invoked and a similar proof yields the conclusion.

Let us now make the dependence of value function on β explicit and use $V_{k,\beta}(\cdot,\cdot)$ to represent the value function in k^{th} step.

Theorem 10: If $c_3 < M$ then for each k and x, $V_{k,\beta}(x, M)$ is non-decreasing in β.

Let $x_k(\beta)$ denote the unique threshold for k-step to go cost function when the discount factor is β.

Lemma 6: The derivative $\frac{d}{dx}V_{k,\beta}(x)$ is non-decreasing function of β.

Theorem 11: The derivative $\frac{d}{d\beta}x_k(\beta) \leq 0$.

C. Average Cost

We now consider the problem of optimization of the infinite horizon average cost when$\mathcal{S}$ is negative. The approach is to use results from infinite horizon discounted cost optimization and then use the standard vanishing discount approach with $\beta \to 1$. It is clear that if the average cost exists, it is independent of the initial state so that we can, without loss of generality, assume that $p_0 = M$. We saw that, for the discounted cost case when $\mathcal{S}$ is negative, once the level M is attained, it is retained throughout. Thus making the structural results obtained in Section IV-B for this particular case very relevant to the analysis of average cost problem. We first need to establish some continuity conditions (conditions **W** in [5]). It is easily shown that the above conditions are satisfied in our problem.

A sufficient condition for existence of stationary average optimal policy, which can be obtained as limit of discounted cost optimal policies $f_\beta(x)$, is provided in [9]. In our problem $f_\beta : \mathcal{R} \to \{0, 1\}$ where 0 means remaining idle and 1 means serving $x \wedge M$ and reordering. Define $w_\beta(x) = V_\beta(x) - \inf_x V_\beta(x)$.

Theorem 12—Schal, Theorem 3.8: Suppose there exists a policy Ψ and an initial state x such that the average cost $V^\Psi(x) < \infty$. Let $\sup_{\beta<1} w_\beta(x) < \infty$ for all x and the Conditions **W** hold, then there exists a stationary policy f_1 which is average cost optimal and the optimal cost is independent of the initial state. Also f_1 is limit discount optimal in the sense that, for any x and given any sequence of discount factors converging to one, there exists a subsequence $\{\beta_m\}$ of discount factors and a sequence $x_m \to x$ such that $f_1(x) = \lim_{m\to\infty} f_{\beta_m}(x_m)$.

In order to apply above result we need to show:

1) Existence of policy Ψ: The policy of serving $x \wedge M$ in every slot yields a finite average cost.
2) $\sup_{\beta<1} w_\beta(x) < \infty$ for all x: For any β, since $V_\beta(x)$ is monotone increasing, it follows that $x^* := \operatorname*{argmin}_x V_\beta(x) = 0$. We have

$$V_\beta(x^*) = h(0) + \beta E_w V_\beta(w).$$

Now, for any fixed $x_0 = x$, consider a policy that serves $x_j \wedge M$ and reorders till the first time the queue is empty (let us denote this time by a random variable Z). Then it can be shown that

$$\begin{aligned} w_\beta(x) &= V_\beta(x) - V_\beta(0) \\ &\leq \sum_{n=\lceil \frac{x}{M} \rceil}^{\infty} [\sum_{j=0}^{n} h(x + \sum_{i=1}^{j} w_i - (j+1)M)] P(Z = n), \end{aligned}$$

where the expression on the right hand side is independent of β and finite almost surely if $E[W] < M$. The required condition is thus verified.

Hence, for the average cost criterion the cost is independent of the initial state. So we can, without loss of generality, assume $p_0 = M$. Now we use the results of Section IV-B to obtain our main result.

Theorem 13: For the average cost optimization problem, if $\mathcal{S}$ is negative and $h(x)$ is a convex function, then there exists a threshold based policy which gives the minimum cost.

V. The Case of $\mathcal{S} = +$ with $B \equiv 0$

We now consider the case when $\mathcal{S}$ is positive and $B \equiv 0$. Our starting point is Theorem 3 which says that if $B \equiv 0$ then the optimal policy serves $x \wedge p$ and then decides whether to reorder the battery or not. Thus, in this case the dynamic programming equation is

$$\begin{aligned} V_{k+1}(x, p) &= h(x) + g(p) \\ &+ \beta \min \Big(E_w V_k((x-p)^+ + w, M) + c_3, \\ &E_w V_k((x-p)^+ + w, (p-x)^+) \Big). \quad (7) \end{aligned}$$

We need to compare $V_k((x-p)^+ + w, M) + c_3$ with $V_k((x-p)^+ + w, (p-x)^+)$ in order to obtain the policy at (x, p). It is noted from these terms that the optimal policy should be a function only of $x - p$, i.e., the decision is same for all (x, p) for which $x - p$ is same. This structure helps us in accurately characterizing the optimal policy which is done below.

A. Finite Horizon Discounted Cost

Consider the dynamic programming equation for the case $x \leq p$.

$$\begin{aligned} V_{k+1}(x, p) &= h(x) + g(p) \\ &+ \beta \min \Big(E_w V_k(w, M) + c_3, E_w V_k(w, p - x) \Big). \quad (8) \end{aligned}$$

Observe that the first term under the minimization operation is independent of x and p. Now, it was shown in Lemma 4 that $\frac{d}{dp}V_k(x, p)$ is bounded above by a negative quantity. Since $V_k(w, M) \leq V_k(w, M) + c_3$, it follows that, for each w, there is a value $p_k^*(w)$ such that $V_k(w, l) > V_k(w, M) + c_3$ for all $l = p_k^*(w)$ and that $V_k(w, l) < V_k(w, M) + c_3$. Note here that it is possible that $p_k^*(w) = 0$ but what is important is that, owing to negative value of $\frac{\partial}{\partial p}V_k(x, p)$, the set $\{l : V_k(w, l) < V_k(w, M) + c_3\}$ is connected and hence has a smallest element that is $p_k^*(w)$. Note also that $p_k^*(w)$ *is independent of the state* (x, p). By taking expectation over w, we obtain

Theorem 14: For a fixed $x < p$, there is a quantity p_k^* such that it is optimal to reorder battery when $x < p < x + p_k^*$ and it is optimal to not reorder battery when $x + p_k^* < p \leq M$.

In order to derive the structure of the optimal policy for the case $x > p$, we again use Lemma 4. Consider the dynamic programming equation for the case $x > p$.

$$V_{k+1}(x,p) = h(x) + g(p)$$
$$+ \beta \min \Big(E_w V_k(x-p+w, M) + c_3, E_w V_k(x-p+w, 0) \Big). \tag{9}$$

¿From Lemma 4, if $\sup_l \frac{d}{dl} g(l) \leq \frac{-c_3}{M}$, then $\frac{d}{dp} V_k(x,p) \leq \frac{-c_3}{M}$ for all values of k thus implying that

$$E_w V_k(x-p+w, M) + c_3 \leq E_w V_k(x-p+w, 0).$$

The condition

$$\sup_l \frac{d}{dl} g(l) < \frac{-c_3}{M}$$

also implies that $p*(w) > 0$ for all values of w, thus $p^* > 0$. Hence we have

Theorem 15: If $\sup_l \frac{d}{dl} g(l) \leq \frac{-c_3}{M}$ then it is optimal to reorder the battery after serving whenever $x > p + p^*$.

This result, along with Theorem 14, gives complete structure of the optimal policy when $\sup_l \frac{d}{dl} g(l) \leq \frac{-c_3}{M}$. Now, note that if it turns out that $\frac{d}{dp} V_k(x,p) > -\frac{c_3}{M}$ for all values of x and p then $p^*(w) = 0$ for all w and that, if $x > p$,

$$V_k(x-p+w, M) + c_3 > V_k(x-p+w, 0) \quad \forall w,$$

so that we get

Theorem 16: If $\inf_p \inf_x \frac{\partial}{\partial x} V_k(x,p) \geq \frac{-c_3}{M}$ for all values of k then it is optimal to never reorder the battery.

The results obtained here are very similar to those obtained for the case when $\mathcal{S}$ is negative in the sense that we get a threshold based policy where existence of a nontrivial threshold depends on the slope of the cost functions. We are now considering the infinite horizon discounted/average cost problems for this case and expect them to provide results of same flavour as those obtained when $\mathcal{S}$ is negative.

VI. Open Loop Control

The key result obtained for the average cost optimization problem was the existence of a threshold based policy for the case when $\mathcal{S}$ is negative. The problem with such an approach is that the threshold depends on the distribution of the arrival process so that the computation of the threshold becomes hard. We may also look at other *suboptimal* policies that are easily implementable. This can be done, for example, by restricting the policies to those that do not require state information. Such policies need not be stationary. A possible decision problem now would be to find an optimal sequence of 0s and 1s where 0 corresponds to reordering a new battery and 1 corresponds to remaining idle. We would like to have a bound on the long term average cost of reordering while minimizing the average buffer occupancy cost.

Let $\{a_n\}_{n \geq 1}$, $a_n \in \{0, 1\}$, be a sequence of controls so that $a_n = 1$ indicates decision of remaining idle at n^{th} decision instant and $a_n = 0$ implies serving the maximum possible amount of fluid and reordering a new battery. We also require an upper bound on the rate at which the battery can be reordered therefore taking in to account the cost of reordering a battery. This can be done by letting

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} a_n \geq p,$$

where p is chosen so that the system is also stable. For the system to be stable, the long term average of given service should satisfy the following inequality.

$$M \cdot \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} (1 - a_n) \geq E[W],$$

i.e., $M(1-p) \geq E[W]$. We now have the problem of minimizing

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} x_n(\mathbf{a})$$

subject to

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} a_n \geq p,$$

where $\mathbf{a}$ is a control sequence and $x_n = (x_{n-1} - (1-a_n)M)^+ + w_n$ is the buffer occupancy at decision instant n. Without loss of generality, we assume that $x_0 = 0$ so that $x_1 = w_1$.

For our case we have that

Theorem 17: The function $f_N(\mathbf{a}) := x_N(\mathbf{a})$ is multimodular for each N.

Now we make the following assumption: The maximum amount of arrival in a slot is bounded by M. Under this assumption, all the conditions in Theorem 6 of [10] are satisfied (we can take the required sequence $b_n \equiv 0$ since $W < M$). We have thus established the optimality of bracket sequences (of rate p) for open loop control of the system under consideration.

VII. Conclusion

We considered jointly optimal scheduling and power control of a wireless device, and formulated it as a Markov decision process problem. We considered the cases of optimizing finite and infinite horizon discounted costs as well as that of infinite horizon average cost. The problem becomes hard as the underlying state space is two-dimensional and important second order properties like convexity or increasing/decreasing differences do not hold. We established the optimality of bang-bang policy which is threshold based. We also studied the behaviour of this threshold and obtained parametric monotonicity results. We then considered the problem of open-loop control of the system where the decision maker does not have knowledge of the system state. For this case we proved that using a bracket sequence based policy results in optimal performance.

ACKNOWLEDGEMENT

This work was supported by project no. 2900-IT-1 from the *Centre Franco-Indien pour la Promotion de la Recherche Avancee* (CEFIPRA).

REFERENCES

[1] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *To appear in IEEE Trans. on Wireless Communications,*, 2004.

[2] W. Ye, J. Heidemann, and D. Estrin, "An Energy-Efficient MAC Protocol for Wireless Sensor Networks," in *Proceedings of the IEEE INFOCOM*, 2002, pp. 1567–1576.

[3] Michele Zorzi and Ramesh R. Rao, "Energy Efficiency of TCP in a Local Wireless Environment," *Mob. Netw. Appl.*, vol. 6, no. 3, pp. 265–278, 2001.

[4] C. E. Price, K. M. Sivalingam, P. Agarwal, and J.-C. Chen, ""A Survey of Energy Efficient Network Protocols for Wireless and Mobile Networks," *ACM/Baltzer Journal on Wireless Networks*, vol. 7, no. 4, pp. 343–358, 2001.

[5] Munish Goyal, Anurag Kumar, and Vinod Sharma, "Power Constrained and Delay Optimal Policies for Scheduling Transmissions over a Fading Channel," in *Proceedings of the IEEE INFOCOM*, 2003.

[6] G. S. Rajadhyaksha and V. S. Borkar, "Transmission Rate Control Over Randomly Varying Channels," *Probability in Engineering and Informational Sciences*, vol. 19, no. 1, pp. 73–82, 2005.

[7] T. F. Fuller, M. Doyle, and J. S. Newman, "Relaxation phenomena in lithium-ion insertion cells," *J. Electrochem. Soc.*, vol. 141, pp. 982–990, 1994.

[8] V. S. Borkar, A. A. Kherani, and B. J. Prabhu, "Control of Buffer and Energy of a Wireless Device: Closed and Open Loop Approaches," Tech. Rep. RR-5414, INRIA, December 2004.

[9] M. Schal, "Average Optimality in Dynamic Programming with General State Space," *Math. of Operations Res.*, vol. 18, pp. 163–172, 1993.

[10] E. Altman, B. Gaujal, and A. Hordijk, *Discrete-event control of stochastic networks: Multimodularity and Regularity*, Lecture Notes in Mathematics. Springer Verlag, 2003.