Optimal Link Adaptation over Partially Observable Gilbert-Elliot Channels

Amine Laourine and Lang Tong
Cornell University, School of Electrical and Computer Engineering, Ithaca, NY
Emails: al496@cornell.edu, ltong@ece.cornell.edu

***Abstract*—In this paper a communication system operating over a Gilbert-Elliot channel is studied. The goal of the transmitter is to maximize the number of successfully transmitted bits. This is achieved by choosing among three possible actions: (i) betting aggressively by using a weak code that allows the transmission of a high number of bits but provides no protection against a bad channel, ii) betting conservatively by using a strong code that perfectly protects the transmitted bits against a bad channel but does not allow a high number of data bits, iii) betting opportunistically by sensing the channel for a fixed duration and then deciding which code to use. The problem is formulated and solved using the theory of Markov decision processes (MDPs). It is shown that the optimal strategy has a simple threshold structure. Closed form expressions and simplified procedures for the computation of the threshold policies in terms of the system parameters are provided.**

I. Introduction

The quality of the radio channel is often random and evolves in time, ranging from good to bad depending on propagation conditions. To cope with this changing behavior and maintain a good quality of service, link adaptation may be performed. Link adaptation is a technique that leads to a better channel utilization by matching the systems parameters of the transmitted signal (data/coding rate, constellation size and transmit power...) to the changing channel conditions [1].

Time-varying fading channels can be well modeled by a finite state Markov chain [2] (and the references therein). A particularly convenient abstraction is the two-state Markovian model known as the Gilbert-Elliot channel [3]. This model assumes that the channel can be in either a good state or a bad state. For example, the channel is in a bad state whenever the SNR drops below a certain threshold and in a good state otherwise.

In this paper we consider a time-slotted communication system operating over a Gilbert-Elliot channel. The transmitter has at its disposal a strong error correcting code and a weak one. The strong code offers perfect protection against the channel errors even if the channel is in a bad state. It however provides the extra protection at the expense of a reduced data rate. The weak code, on the other hand, offers perfect protection against the channel errors when the channel is in the good state but fails otherwise. At the beginning of each time slot, the transmitter can choose among three possible actions: i) transmitting at a low data rate using the strong error correcting code, ii) transmitting at a high data rate using the weak error correcting code, and iii) sensing the channel for a fraction of the slot and then use the appropriate code. The extra knowledge provided by this last action comes at a price, which is the time spent probing the channel. We take as objective the maximization of the total expected discounted number of bits transmitted over an infinite time span. We formulate and solve the problem using Markov decision processes (MDP).

MDP tools have been previously applied to solve communication problems over time-varying channels. Most related to this paper are [4] and [5]. In [4], the authors employed results from optimal search theory and provided threshold strategies that minimize the transmission energy and delay associated with transmitting a file over a Gilbert-Elliot channel. Similarly in [5], taking as objective the maximization of the throughput and the minimization of the energy consumption, the authors established the optimality of the threshold policies. The effect of the sensing action on the throughput of a communication system was not considered in these papers.

A closely related area to the problem studied here is the so-called opportunistic spectrum access (refer to [6] for an overview) where sensing is an integral part of the access scheme. A generic setup is as follows: a secondary user tries to opportunistically access a channel which, depending on the state of the primary user, can be either busy or idle. The problem considered here is different in that the transmitter is allowed to transmit without first probing the channel. In addition, we model explicitly the cost of sensing. Thus, the sensing action must be judiciously used in order to maximize the total number of transmitted bits.

The technique used in this paper has its origin in [7], where Ross considered the problem of quality control of a production process modeled by a special two-state Markov chain. Specialized for wireless transmissions, our model is different in that the good and bad states of the channel are independent from the action of the user. However, in Ross's paper, the bad state of the production process can only change back to the good state under the revise action. This fact, renders the immediate application of Ross's results nontrivial. The problem at hand therefore deserves a proper theoretical treatment.

The rest of the paper is organized as follows, in Section II we formulate the problem as a Markov decision process. In Section III, we use methods developed in the context of quality control and reliability theory [7]-[9] to establish the optimality

This work is supported in part by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program DAAD19-01-2-0011.

of threshold policies. In Section IV, we provide closed form expressions and simplified procedures for the computation of the thresholds in terms of the system parameters. In Section V, we also provide closed form expressions of the optimal total expected discounted number of bits transmitted. In Section VI, we provide numerical examples to illustrate the various theoretical results that will be presented in the paper. Finally, Section VII concludes the paper.

II. Problem formulation

A. Channel model and assumptions

We consider a communication system operating over a slotted Gilbert-Elliot channel which is a one dimensional Markov chain G_n with two states: a good state denoted by 1 and a bad state denoted by 0. The channel transition probabilities are given by $\Pr[G_n = 1|G_{n-1} = 1] = \lambda_1$ and $\Pr[G_n = 1|G_{n-1} = 0] = \lambda_0$. We assume that $\lambda_0 \leq \lambda_1$, the so-called positive correlation assumption, which can be restrictive in practice though it simplifies the analysis considerably (similar assumption have also been used in [4], [5]). From now on we assume without loss of generality that the slot duration is a unity, so that we will interchangeably use data rate and number of bits.

B. Communication protocol

At the beginning of each slot, the transmitter can choose among three possible actions: betting conservatively, betting aggressively, and betting opportunistically.

Betting conservatively: For this action (denoted by T_l), the transmitter decides to "play safe" and transmits a low number R_1 of data bits. This corresponds to the situation when the transmitter believes that the channel is in a bad state. Hence the transmitter uses a strong error correcting code with a high redundancy thereby leading to the transmission of a smaller number of data bits. If this action is chosen, we assume that the transmission is successful regardless of the channel quality. Hence, in this situation, the receiver is not required to reply back with an ACK, since the transmitter is assured that the transmission was successful.

Betting aggressively: For this action (denoted by T_h), the transmitter decides to "gamble" and transmits a high number R_2 ($> R_1$) of data bits. This corresponds to the situation when the transmitter believes that the channel is in a good state. If this action is taken we assume that the transmission is successful only if the channel is in the good state. At the end of the slot, the transmitter will receive an ACK if the channel was in the good state, and will receive a NAK otherwise. Hence, if this action is chosen, the transmitter will learn the channel state during the elapsed slot.

Betting opportunistically: For this action (denoted by S), the transmitter decides to sense the channel at the beginning of the slot. We assume that sensing is perfect, *i.e.*, sensing reveals the true state of the channel. We assume also that sensing lasts a fraction $\tau (< 1)$ of the slot. Sensing can be carried out by making the transmitter send a control/probing packet. Then, the receiver responds with a packet indicating the channel state. Depending on the sensing outcome, the transmitter will transmit $(1-\tau)R_1$ data bits if the channel was found to be in the bad state or $(1-\tau)R_2$ data bits if otherwise. This extra knowledge comes at a price, which is the time spent probing the channel. However, the sensing action offers the advantage of updating the belief (the posterior estimate) about the channel state. This updated belief can be exploited in the future slots in order to increase the throughput. This fact captures a fundamental tradeoff known as the exploration-exploitation dilemma. Note finally that in this situation the receiver is not required to reply back with an ACK, since the transmitter is assured that the transmission was successful.

C. MDP formulation

At the beginning of a time slot, the transmitter is confronted with a choice among three actions. It must judiciously select actions so as to maximize a ceratin reward to be defined shortly. Because the state of the channel is not directly observable, the problem in hand is a Partially Observable Markov Decision Process (POMDP). In [10], it is shown that a sufficient statistic for determining the optimal policy is the conditional probability that the channel is in the good state at the beginning of the current slot given the past history (henceforth called belief) denoted by $X_t = \Pr[G_t = 1|\mathcal{H}_t]$, where $\mathcal{H}_t$ is all the history of actions and observations at the current slot t. Hence by using this belief as the decision variable, the POMDP problem is converted into an MDP with the uncountable state space $[0, 1]$.

Define a policy π as a rule that dictates the action to choose, *i.e.*, a map from the belief at a particular time to an action in the action space. Let $V_\beta^\pi(p)$ be the expected discounted reward with initial belief $X_0 = \Pr[G_0 = 1|\mathcal{H}_0] = p$, where the superscript π denotes the policy being followed and the subscript β ($\in [0, 1)$) the discount factor. The expected discounted cost has the following expression

$$V_\beta^\pi(p) = \mathbb{E}^\pi\left[\sum_{t=0}^{\infty} \beta^t R(X_t, A_t)|X_0 = p\right], \qquad (1)$$

where t is the time slot index, A_t is the action chosen at time t, $A_t \in \{T_l, S, T_h\}$. The term $R(X_t, A_t)$ denotes the reward acquired when the belief is X_t and the action A_t is chosen:

$$R(X_t, A_t) = \begin{cases} R_1 & \text{if } A_t = T_l \\ (1-\tau)[(1-X_t)R_1 + X_t R_2] & \text{if } A_t = S \\ X_t R_2 & \text{if } A_t = T_h \end{cases}.$$

These equations can be explained as follows: when betting conservatively, R_1 bits are transmitted regardless of the channel conditions and the transmission is always successful. When betting aggressively, R_2 bits are transmitted if the channel happens to be in the good state whereas 0 bits are transmitted if the channel was in the bad state. Hence, since the belief that the channel is in the good state is X_t, the expected return when the risky action is taken is $X_t R_2$. Now, when the sensing action is taken $(1-\tau)R_1$ bits will be transmitted if the sensing revealed that the channel was in a bad state whereas $(1-\tau)R_2$

bits will be transmitted otherwise. Hence the expected return when the sensing action is taken is $(1-\tau)[(1-X_t)R_1+X_tR_2]$.

Define now the value function $V_\beta(p)$ as

$$V_\beta(p) = \max_\pi V_\beta^\pi(p) \quad \text{for all } p \in [0,1]. \tag{2}$$

A policy is said to be stationary if it is a function mapping the state space $[0,1]$ into the action space $\{T_l, S, T_h\}$. It is well known [11, Th.6.3] that there exists a stationary policy π^* such that $V_\beta(p) = V_\beta^{\pi^*}(p)$. The value function $V_\beta(p)$ satisfies the Bellman equation

$$V_\beta(p) = \max_{A\in\{T_l,S,T_h\}}\{V_{\beta,A}(p)\}, \tag{3}$$

where $V_{\beta,A}(p)$ is the value acquired by taking action A when the initial belief is p and is given by

$$V_{\beta,A}(p) = R(p,A) + \beta\mathbb{E}_Y[V_\beta(Y)|X_0 = p, A_0 = A], \tag{4}$$

where Y denotes the next belief when the action A is chosen and the initial belief is p. The term $V_{\beta,A}(p)$ will be explained next for the three possible actions.

a) Betting conservatively: If this action is taken, R_1 bits will be successfully transmitted regardless of the channel quality. The transmitter will not learn what was the channel quality. Hence, if the transmitter had a belief p during the elapsed time slot, its belief at the beginning of the next time slot is given by

$$T(p) = \lambda_0(1-p) + \lambda_1 p = \alpha p + \lambda_0, \tag{5}$$

with $\alpha = \lambda_1 - \lambda_0$. Consequently $V_{\beta,T_l}(p)$ is given by

$$V_{\beta,T_l}(p) = R_1 + \beta V_\beta(T(p)). \tag{6}$$

b) Betting opportunistically: If this action is taken and the current belief is p, the channel quality during the current slot is then revealed to the transmitter. With probability p the channel will be in the good state and hence the belief at the beginning of the next slot will be λ_1. Likewise, with probability $1-p$ the channel will turn out to be in the bad state and hence the updated belief for the next slot is λ_0. Consequently $V_{\beta,S}(p)$ is given by

$$V_{\beta,S}(p)=(1-\tau)[pR_2+(1-p)R_1]+\beta[pV_\beta(\lambda_1)+(1-p)V_\beta(\lambda_0)]. \tag{7}$$

c) Betting aggressively: If this action is taken and the current belief is p, then with probability p, the transmission will be successful and the transmitter will receive an ACK from the receiver. The belief at the beginning of the next slot will be then λ_1. Similarly, with probability $1-p$, the channel will turn out to be in the bad state and the transmission will result in a failure accompanied by a NAK from the receiver. Hence the transmitter will update his belief for the next slot to λ_0. Consequently $V_{\beta,T_h}(p)$ is given by

$$V_{\beta,T_h}(p) = pR_2 + \beta[pV_\beta(\lambda_1) + (1-p)V_\beta(\lambda_0)]. \tag{8}$$

Finally the Bellman equation for our communication problem reads as follows

$$V_\beta(p) = \max\{V_{\beta,T_l}(p), V_{\beta,S}(p), V_{\beta,T_h}(p)\}. \tag{9}$$

III. Structure of the Optimal Policy

In the following, we will prove the optimality of the threshold policies. But before we need to prove some results about the value function.

Theorem 1. *$V_\beta(p)$ is convex and nondecreasing.*

Proof: See [12]. ■

Using the convexity of $V_\beta(p)$, we are now ready to characterize the structure of the optimal policy.

Theorem 2. *Let $p \in [0,1]$, there are numbers $0 \le \rho_1 \le \rho_2 \le \rho_3 \le 1$ such that*

$$\pi^*(p) = \begin{cases} T_l & \text{if } 0 \le p < \rho_1 \text{ or } \rho_2 < p < \rho_3 \\ S & \text{if } \rho_1 \le p \le \rho_2 \\ T_h & \text{if } \rho_3 \le p \le 1 \end{cases}.$$

Proof: We introduce the following sets

$$\Phi_\mathcal{K} = \{p \in [0,1], V_\beta(p) = V_{\beta,\mathcal{K}}(p)\}, \quad \mathcal{K} \in \{T_l, T_h, S\}. \tag{10}$$

In other words, $\Phi_\mathcal{K}$ is the set of beliefs for which it is optimal to take the action $\mathcal{K}$. The proof uses the convexity of $V_\beta(p)$ in order to show that Φ_{T_h} and Φ_S are convex sets. Since convex subsets of the real line are intervals and $1 \in \Phi_{T_h}$, then there exists $\rho_3 \in (0,1]$ such that $\Phi_{T_h} = [\rho_3, 1]$. Similarly, there exits $\rho_1, \rho_2 \in [0,1]$ such that $\Phi_S = [\rho_1, \rho_2]$. Whence we have that $\Phi_{T_l} = [0, \rho_1) \cup (\rho_2, \rho_3)$. For further details, see [12]. ■

The established structure is appealing since the belief space is partitioned into at most 4 regions. Intuitively, one would think that there should exist only three regions, *i.e.*, if the belief is small, one should paly safe; if the belief is high, one should gamble, and somewhere in between sensing is optimal. Therefore it may seem possible that $(\rho_2, \rho_3) = \emptyset$. However, we show in Section VI that this is not true in general, for some cases, a three-threshold policy is optimal.

IV. Closed form characterization of the policies

Theorem 2. proves that there exists three types of threshold policies; a one-threshold policy (when $\rho_1 = \rho_2 = \rho_3$), a two-thresholds policy (when $\rho_1 < \rho_2 = \rho_3$), and a three-thresholds policy (when $\rho_1 < \rho_2 < \rho_3$). Since we do not have sufficient and necessary conditions to tell which policy will be optimal, one will need to compute the three possible policies and select the one that achieves the highest value. Fortunately, this computation is inexpensive because we will provide closed form expressions and simplified procedures to compute the policies. Also, depending on the system parameters, some policies may be infeasible. For example, in a 2-thresholds policy, we would find $\rho_1 > \rho_2$. In such situations, the task is even more simplified since we can further restrict our search for the optimal policy.

In the following we will analyze each policy individually, but before delving into the computation of the thresholds, we need to introduce the following operators:

$$T^n(p) = T(T^{n-1}(p)) = \lambda_F(1-\alpha^n) + \alpha^n p. \tag{11}$$

$$T^{-n}(p) = T^{-1}(T^{-(n-1)}(p)) = \frac{p}{\alpha^n} - \frac{1-\alpha^n}{1-\alpha}\frac{\lambda_0}{\alpha^n}. \quad (12)$$

We will denote also by $\lambda_F = \frac{\lambda_0}{1-\alpha}$ the fixed point of $T(\cdot)$, *i.e.*, $T(\lambda_F) = \lambda_F$ c.f. (5).

A. One threshold policy

Assume that the optimal policy has one threshold $0 < \rho < 1$. The procedure to calculate ρ starts by computing $V_\beta(\lambda_0)$ and $V_\beta(\lambda_1)$ as shown in section A in the Appendix. The threshold ρ is computed as in the following lemma.

Lemma 1. If the one threshold policy is optimal then the threshold ρ is calculated as follows:
If[1] $\frac{R_1}{1-\beta} \geq V_{\beta,T_h}(\lambda_F)$, then

$$\rho = \frac{R_1}{R_2 + \beta V_\beta(\lambda_1) - \beta\frac{R_1}{1-\beta}}. \quad (13)$$

Otherwise, we have

$$\rho = \frac{(1-\beta\lambda_1)R_1 + \beta\lambda_0 R_2 + \beta(\beta-1)(1-\beta\alpha)V_\beta(\lambda_0)}{(1-\beta\alpha)(R_2 + \beta(\beta-1)V_\beta(\lambda_0))}. \quad (14)$$

Proof: See [12]. ■

B. Two thresholds policy

Assume that the optimal policy has two thresholds $0 < \rho_1 < \rho_2 < 1$. Note that since ρ_2 is the solution of $V_{\beta,S}(\rho_2) = V_{\beta,T_h}(\rho_2)$, it is easy to establish that $\rho_2 = \frac{(1-\tau)R_1}{(1-\tau)R_1+\tau R_2}$. The procedure to compute ρ_1 starts by computing $V_\beta(\lambda_0)$ and $V_\beta(\lambda_1)$ as in section B in the Appendix. The threshold ρ_1 is computed as in the following lemma.

Lemma 2. If the two-thresholds policy is optimal then ρ_1 is computed as follows
1) If $\lambda_F > \rho_2$ then two cases can be distinguished:
If[2] $V_{\beta,T_l}(T^{-1}(\rho_2)) < V_{\beta,S}(T^{-1}(\rho_2))$, ρ_1 will be equal to (15) given at the bottom of this page.
Else ρ_1 will be equal to (16) given at the bottom of this page.
2) If $\frac{R_1}{1-\beta} < V_{\beta,S}(\lambda_F)$ and $\lambda_F \leq \rho_2$, ρ_1 is given by (15).
3) Finally if $\frac{R_1}{1-\beta} \geq V_{\beta,S}(\lambda_F)$ and $\lambda_F \leq \rho_2$, then

$$\rho_1 = \frac{\tau(1-\beta)R_1}{(1-\tau)(1-\beta)(R_2-R_1) + \beta((1-\beta)V_\beta(\lambda_1) - R_1)}. \quad (17)$$

Proof: See [12]. ■

C. Three thresholds policy

Assume that the optimal policy has three thresholds $0 < \rho_1 < \rho_2 < \rho_3 < 1$. Before detailing the structure of the optimal policy, we introduce the following useful lemma.

Lemma 3. If the three-thresholds policy is optimal, then $\lambda_F \in [\rho_3, 1]$.

Proof: See [12]. ■

We now turn to the computation of ρ_1, ρ_2 and ρ_3. Since $\rho_3 \leq \lambda_F$ and ρ_3 is the solution of $R_1 + \beta V_\beta(T(\rho_3)) = V_{\beta,T_h}(\rho_3)$, it follows that ρ_3 is given by (14). The two other thresholds ρ_1 and ρ_2 are computed as in the following lemma.

Lemma 4. If the three thresholds policy is optimal then let $J+1 = \min\{k \geq 1 : \delta(k) < \gamma(k)\rho_3\}$, where $\gamma(k)$ is given by

$$\begin{aligned}\gamma(k) &= [(1-\tau)(R_2-R_1) + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0))]\frac{1}{\alpha^k} \\ &- \beta^k[R_2 + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0))], \end{aligned} \quad (18)$$

and

$$\begin{aligned}\delta(k) &= R_1\frac{1-\beta^k}{1-\beta} + \beta(\beta^k - 1)V_\beta(\lambda_0) - (1-\tau)R_1 \\ &+ \frac{\lambda_0(1-\alpha^k)}{\alpha^k(1-\alpha)}[(1-\tau)(R_2-R_1) + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0))]. \end{aligned} \quad (19)$$

We have then that ρ_2 is equal to (20) given at the bottom of this page.
If $V_{\beta,T_l}(T^{-1}(\rho_2)) < V_{\beta,S}(T^{-1}(\rho_2))$ then ρ_1 will be given by (15). Else, let $J'+1 = \min\{k \geq J+2 : \gamma(k)\rho_3 < \delta(k)\}$, and ρ_1 will be given by (20) with J replaced by J'.

Proof: See [12]. ■

V. Computation of the value function

Since $V_{\beta,S}(p)$ and $V_{\beta,T_h}(p)$ are linear functions of p, once $V_\beta(\lambda_0)$ and $V_\beta(\lambda_1)$ are computed, $V_\beta(p)$ is completely determined when $p \in \Phi_S \bigcup \Phi_{T_h}$. $V_\beta(p)$ needs however to be determined for $p \in \Phi_{T_l}$.

[1] Note that $V_{\beta,T_h}(\lambda_F)$ is directly computable since we have calculated $V_\beta(\lambda_0)$ and $V_\beta(\lambda_1)$ in the previous step.

[2] Note that $V_{\beta,T_l}(T^{-1}(\rho_2)) = R_1 + \beta([R_2 + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0))]\rho_2 + \beta V_\beta(\lambda_0))$ is readily computable since we have already calculated $V_\beta(\lambda_0)$ and $V_\beta(\lambda_1)$. The same remark holds for $V_{\beta,S}(T^{-1}(\rho_2))$.

$$\rho_1 = \frac{\tau R_1 + \beta(1-\tau)[R_1 + \lambda_0(R_2-R_1)] + \beta^2[V_\beta(\lambda_0) + \lambda_0(V_\beta(\lambda_1) - V_\beta(\lambda_0))] - \beta V_\beta(\lambda_0)}{(1-\beta\alpha)[(1-\tau)(R_2-R_1) + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0))]} \quad (15)$$

$$\rho_1 = \frac{\beta\lambda_0 R_2 + (1-\beta\lambda_1)\tau R_1 + \beta(\beta-1)(1-\beta\alpha)V_\beta(\lambda_0)}{R_2(1-\lambda_1(\beta(\alpha-\tau)-\tau)) + \beta(\beta-1)(1-\beta\alpha)V_\beta(\lambda_0) - (1-\tau)(1-\beta\lambda_1)R_1} \quad (16)$$

$$\rho_2 = \frac{R_1(\frac{1-\beta^{J+1}}{1-\beta} - (1-\tau)) + \beta^{J+1}[P_F(1-\alpha^{J+1})(R_2 + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0))) + \beta V_\beta(\lambda_0)] - \beta V_\beta(\lambda_0)}{(1-\tau)(R_2-R_1) + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0)) - (\alpha\beta)^{J+1}(R_2 + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0)))} \quad (20)$$

A. One threshold policy

The goal is to find $V_\beta(p)$ for $p \le \rho$. Here we can distinguish two possibilities
If $\lambda_F \le \rho$, then $V_\beta(p) = \frac{R_1}{1-\beta}$ for all $p \le \rho$ (see lemma 5 in the appendix). If $\lambda_F > \rho$, then let $J{+}1 = \min\{k \in \mathbb{N} : T^{-k}(\rho) < 0\}$. Let $F_{J+1} = [0, T^{-J}(\rho)]$ and $F_i = (T^{-i}(\rho), T^{-(i-1)}(\rho)]$ for $1 \le i \le J$. Then for $p \in F_i$, we have $T^i(p) > \rho \ge T^{(i-1)}(p)$, *i.e.*,

$$V_\beta(p) = R_1 \frac{1-\beta^i}{1-\beta} + \beta^i V_{\beta,T_h}(T^i(p)). \tag{21}$$

The optimal policy for this last case is illustrated in Fig. 1.

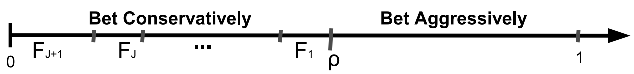

Fig. 1. Illustration of the one threshold policy for $\lambda_F > \rho$.

B. Two thresholds policy

The approach here is similar to the previous case, *i.e.*,
If $\lambda_F \le \rho_1$, then $V_\beta(p) = \frac{R_1}{1-\beta}$ for all $p \le \rho_1$. If $\rho_1 < \lambda_F \le \rho_2$, let $J{+}1 = \min\{k \in \mathbb{N} : T^{-k}(\rho_1) < 0\}$. Let $F_{J+1} = [0, T^{-J}(\rho_1)]$ and $F_i = (T^{-i}(\rho_1), T^{-(i-1)}(\rho_1)]$ for $1 \le i \le J$. Then for $p \in F_i$, we have $\rho_2 > T^i(p) > \rho_1 \ge T^{(i-1)}(p)$, *i.e.*,

$$V_\beta(p) = R_1 \frac{1-\beta^i}{1-\beta} + \beta^i V_{\beta,S}(T^i(p)). \tag{22}$$

The optimal policy for this case is illustrated in Fig. 2.

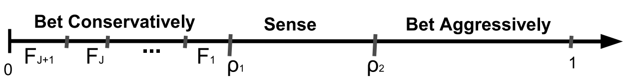

Fig. 2. Illustration of the two thresholds policy for $\rho_1 < \lambda_F \le \rho_2$.

If $\lambda_F > \rho_2$, two cases can be distinguished; If $T(\rho_1) \le \rho_2$ then the computation is similar to the situation where $\rho_1 < \lambda_F \le \rho_2$ discussed above. If $T(\rho_1) > \rho_2$, let $F_{J+1} = [0, T^{-J}(\rho_1)]$, for $2 \le i \le J$ let $F_i = (T^{-i}(\rho_1), T^{-(i-1)}(\rho_1)]$ and $F_1 = (T^{-1}(\rho_1), T^{-1}(\rho_2)]$. Then for $p \in F_i, i \ge 1$, $V_\beta(p)$ will be given by (22). For $p \in F_0 = (T^{-1}(\rho_2), \rho_1]$ we have

$$V_\beta(p) = R_1 + \beta V_{\beta,T_h}(T(p)). \tag{23}$$

C. Three thresholds policy

The goal is to find $V_\beta(p)$ for $p \in [0,\rho_1] \cup [\rho_2, \rho_3]$. For $p \in [\rho_2, \rho_3]$, let $J+1 = \min\{k \in \mathbb{N} : T^{-k}(\rho_3) < \rho_2\}$. Let $F_{J+1} = [\rho_2, T^{-J}(\rho_3))$ and $F_i = [T^{-i}(\rho_3), T^{-(i-1)}(\rho_3))$ for $1 \le i \le J$. For $p \in F_i$, we have $T^i(p) \ge \rho_3$, *i.e.*, $V_\beta(p)$ is given by (21).
For $p \in [0, \rho_1]$ we can distinguish two cases;
If $T(\rho_1) \le \rho_2$, $V_\beta(p)$ for $p \in [0,\rho_1]$ is computed using (22).
If $T(\rho_1) > \rho_2$, let $H{+}1 = \min\{k \in \mathbb{N} : T^{-k}(\rho_1) < 0\}$. Then we have two subcases: If $T^{-(H+1)}(\rho_2) \ge 0$, then let $Z_{H+1} = [0, T^{-(H+1)}(\rho_2))$, for $1 \le i \le H$ let $Z_i = [T^{-i}(\rho_1), T^{-i}(\rho_2))$ and for $1 \le i \le H+1$ let $Q_i = [T^{-i}(\rho_2), T^{-(i-1)}(\rho_1))$. For $p \in Z_i$, $T^i(p) \in [\rho_1, \rho_2)$ and hence $V_\beta(p)$ is computed using (22). For $p \in Q_i$, $T^i(p) \in [\rho_2, \rho_3)$, hence there exits $1 \le j \le J+1$ such that $T^i(p) \in F_j$, *i.e.*,

$$V_\beta(p) = R_1 \frac{1-\beta^{i+j}}{1-\beta} + \beta^{i+j} V_{\beta,T_h}(T^{i+j}(p)). \tag{24}$$

The optimal policy for this case is illustrated in Fig. 3.

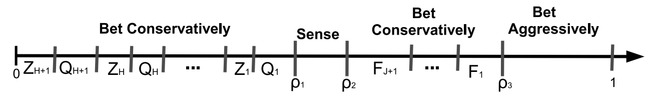

Fig. 3. Illustration of the three thresholds policy for $T(\rho_1) > \rho_2$ and $T^{-(H+1)}(\rho_2) \ge 0$.

If $T^{-(H+1)}(\rho_2) < 0$, then let $Z_{H+1} = [0, T^{-H}(\rho_1))$, for $1 \le i \le H$ let $Z_i = [T^{-i}(\rho_2), T^{-(i-1)}(\rho_1))$ and $Q_i = [T^{-i}(\rho_1), T^{-i}(\rho_2))$. For $p \in Z_i$, $T^i(p) \in [\rho_2, \rho_3)$ and hence $V_\beta(p)$ is given by (24). For $p \in Q_i$, $T^i(p) \in [\rho_1, \rho_2)$ and consequently $V_\beta(p)$ is computed using (22).

VI. Numerical results

We will first consider three different scenarios each one of them leading to a different optimal policy. To validate the closed-form solutions obtained above we will also generate the optimal value function $V_\beta(p)$ using the value iteration algorithm.
The parameters chosen below are selected in order to illustrate that in theory, any of the three policies could be optimal. The first set of parameters considered is $\lambda_0 = 0.2$, $\lambda_1 = 0.9$, $\tau = 0.4$, $R_1 = 1$, $R_2 = 2$ and $\beta = 0.1$. Note that from a practical standpoint $\tau = 0.4$ represents a substantial duration for sensing.
As shown in Fig. 4, the optimal policy in this case is a one threshold policy, whereas the two and three thresholds policies are unfeasible in this case. If we keep all the parameter values fixed and diminish the sensing time to $\tau = 0.1$, then from Fig.5, we can see that the optimal policy becomes a two thresholds policy, whereas the one threshold policy gives suboptimal values (the three thresholds policy is unfeasible in this case).
Fig. 6 shows the optimal value function for the following settings: $\lambda_0 = 0.81$, $\lambda_1 = 0.98$, $\tau = 0.035$, $R_1 = 2.91$, $R_2 = 3$ and $\beta = 0.7$. Here, the optimal policy is a three thresholds policy, and the one and two thresholds policies provide suboptimal results. These numerical simulations prove that all scenarios can be possible and that our developed formulae give always the optimal policy. Finally, it should be noted that finding a scenario where the optimal policy has three-thresholds was not obvious. The parameters had to be repeatedly tuned in order to obtain such a case.
Fig. 7 shows the effect of the sensing time τ on the length of the sensing region $|\Phi_S| = \rho_2 - \rho_1$. The system parameters in this plot are as follows: $R_1 = 1$, $R_2 = 2$, $\beta = 0.99$, $\lambda_0 = 0.1$ and $\lambda_1 = 0.9$. In this example, the two-thresholds policy is optimal for $\tau \in [0, 0.537]$, and beyond this critical value,

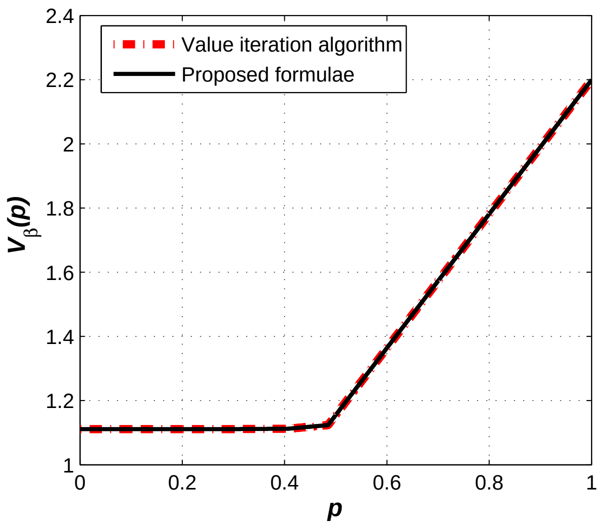

Fig. 4. Optimality of a one threshold policy.

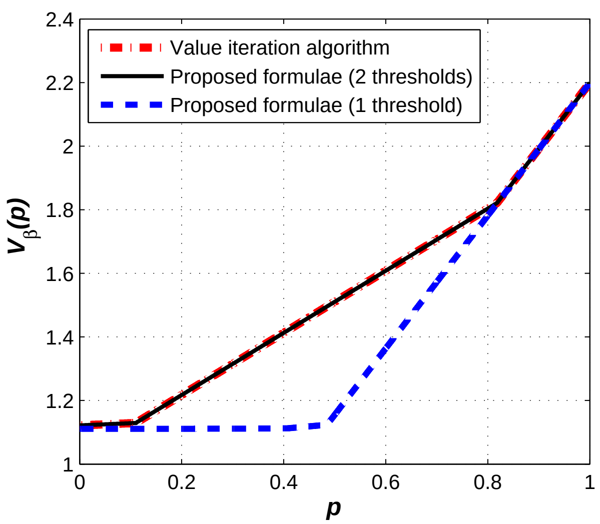

Fig. 5. Optimality of a two thresholds policy.

the one threshold policy will become optimal. As expected, the sensing region Φ_S expands when the cost of sensing τ decreases until it covers the whole interval $[0,1]$ when $\tau = 0$.

VII. Conclusion

In this paper, we have studied a communication system operating over a Gilbert-Elliot channel. In order to maximize the number of successfully transmitted bits, the transmitter judiciously selects the best action among three possible options: i) transmit a high number of bits with no protection against a bad channel, ii) transmit a low number of bits but with perfect protection, iii) sense the channel for a fixed duration and then decide between the two previous actions.

We have formulated the aforementioned problem as a Markov Decision Process, and we have established that the optimal strategy is a threshold policy. Namely, we have proved

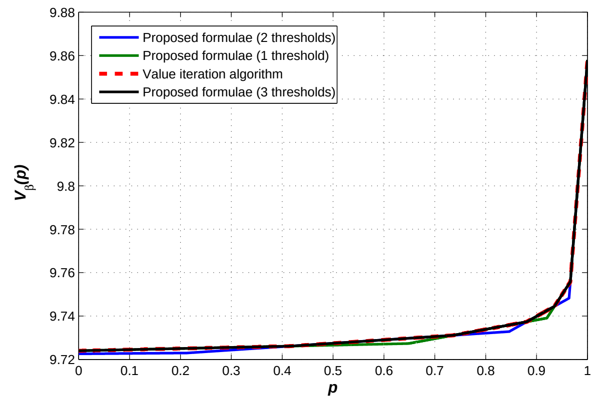

Fig. 6. Optimality of a three thresholds policy.

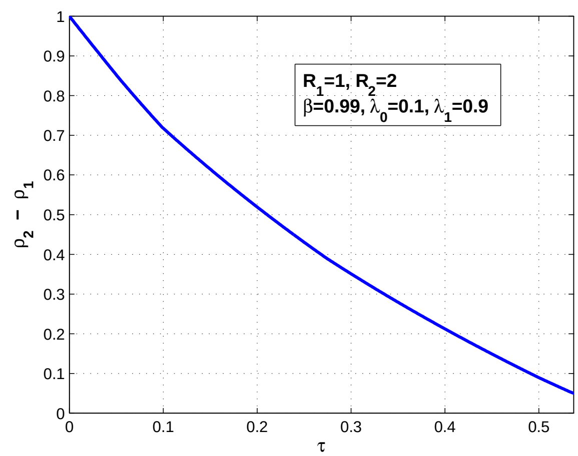

Fig. 7. The effect of the sensing duration τ on Φ_S.

that the optimal policy can have either one threshold, two thresholds, or three thresholds. We have provided closed-form expressions and simplified procedures for the computation of these policies as well as the resulting optimal number of transmitted bits. From a practical standpoint, the results presented in this paper could be used to optimize the channel utilization of real systems such as High-Speed Downlink Packet Access (HSDPA).

Appendix: Computation of $V_\beta(\lambda_1)$ and $V_\beta(\lambda_0)$

Before giving the expressions of $V_\beta(\lambda_1)$ and $V_\beta(\lambda_0)$ we present an alternate expression for $V_\beta(p)$. This new expression will prove to be useful in the subsequent derivations.

Theorem 3. *The value function can be written as*

$$V_\beta(p)=\max_{n\geq 0}\left\{\frac{1-\beta^n}{1-\beta}R_1+\beta^n \max\{V_{\beta,S}(T^n(p)), V_{\beta,T_h}(T^n(p))\}\right\}. \quad (25)$$

Proof: See [12]. ∎

Intuitively the previous result can be explained as follows; The expression $\frac{1-\beta^n}{1-\beta}R_1 + \beta^n V_{\beta,S}(T^n(p))$ is the expected return when the transmitter selects n (≥ 0) times the action T_l, then selects the action S and the procedure continues on there on optimality. Similarly for the other term but instead of taking the S action at the $(n+1)$th stage, the action T_h is selected. The value function is then just the maximum between these two expressions over all stages.

Before proceeding with the computation of $V_\beta(\lambda_1)$ and $V_\beta(\lambda_0)$ we need the following lemma.

Lemma 5. For the one and two-thresholds policies, let $\Phi_{T_l} = [0, \rho)$. If $\lambda_F \in \Phi_{T_l}$ then $V_\beta(p) = \frac{R_1}{1-\beta}$ for all $p \in \Phi_{T_l}$.

Proof: For all $p \leq \lambda_F$, $V_\beta(p) = R_1 + \beta V_\beta(T(p))$, however, $p \leq T(p) \leq \lambda_F$, hence $V_\beta(T(p)) = R_1 + \beta V_\beta(T^2(p))$, *i.e.*, $V_\beta(p) = R_1(1+\beta) + \beta^2 V_\beta(T^2(p))$. By induction we obtain

$$V_\beta(p) = R_1 \frac{1-\beta^n}{1-\beta} + \beta^n V_\beta(T^n(p)) \quad \text{for all } n. \tag{26}$$

We obtain the desired result by letting $n \to \infty$ (since $0 \leq \beta < 1$). Similarly, for $\lambda_F \leq p \leq \rho$, $V_\beta(p) = R_1 + \beta V_\beta(T(p))$, however, $p \geq T(p) \geq \lambda_F$, hence by induction we arrive at the same conclusion. ■

We are now ready to compute $V_\beta(\lambda_1)$ and $V_\beta(\lambda_0)$ for each policy individually.

A. One threshold policy

There are two possible scenarios:

If $\lambda_1 \leq \rho$ then since $\lambda_F \leq \lambda_1 \leq \rho$, from lemma 5, we have $V_\beta(\lambda_1) = V_\beta(\lambda_0) = \frac{R_1}{1-\beta}$.

If $\lambda_1 > \rho$ then $V_\beta(\lambda_1) = V_{\beta,T_h}(\lambda_1)$, *i.e.*, $V_\beta(\lambda_1) = \frac{\lambda_1 R_2 + \beta(1-\lambda_1)V_\beta(\lambda_0)}{1-\beta\lambda_1}$ and using (25), we have that $V_\beta(\lambda_0)$ is a solution to the following equation

$$\begin{aligned} V_\beta(\lambda_0) &= \max_{n\geq 0}\left\{\frac{1-\beta^n}{1-\beta}R_1 + \beta^n V_{\beta,T_h}(T^n(\lambda_0))\right\} \\ &= \max_{n\geq 0}\{\frac{1-\beta^n}{1-\beta}R_1 + \beta^n(\kappa_n R_2 + \beta(V_\beta(\lambda_0) \\ &+ \kappa_n(V_\beta(\lambda_1) - V_\beta(\lambda_0)))))\}, \end{aligned} \tag{27}$$

where $\kappa_n = T^n(\lambda_0) = (1-\alpha^{n+1})\lambda_F$. Hence solving for $V_\beta(\lambda_0)$ we obtain

$$V_\beta(\lambda_0) = \max_{n\geq 0}\left\{\frac{\frac{1-\beta^n}{1-\beta}R_1 + \beta^n g_n R_2}{1-\beta^{n+1}[1-(1-\beta)g_n]}\right\}, \tag{28}$$

where $g_n = \frac{\kappa_n}{1-\beta\lambda_1}$. Note that the last maximization is just a one dimensional search and is computationally inexpensive. Indeed, since $0 \leq \beta < 1$, the search for a maximum can be effectively restricted to values of $n \leq N$, where N is a sufficiently large value such that $\beta^N \ll 1$.

Once $V_\beta(\lambda_0)$ and $V_\beta(\lambda_1)$ have been computed for both cases, we retain the scenario that achieves the maximal values. Indeed, from (2), it is seen that the optimal policy is the one that gives the maximal value for any initial belief p. Hence, in particular, the threshold ρ should be tuned so as to maximize $V_\beta(\lambda_0)$ and $V_\beta(\lambda_1)$.

B. Two thresholds policy

There are three possible scenarios:

If $\lambda_1 \leq \rho_1$ then $V_\beta(\lambda_1) = V_\beta(\lambda_0) = \frac{R_1}{1-\beta}$. If $\rho_1 \leq \lambda_1 \leq \rho_2$ then $V_\beta(\lambda_1) = V_{\beta,S}(\lambda_1)$, *i.e.*, $V_\beta(\lambda_1) = \frac{(1-\tau)[R_1+\lambda_1(R_2-R_1)]+\beta(1-\lambda_1)V_\beta(\lambda_0)}{1-\beta\lambda_1}$. Hence, using (25) we have $V_\beta(\lambda_0) = \max_{n\geq 0}\left\{\frac{1-\beta^n}{1-\beta}R_1 + \beta^n V_{\beta,S}(T^n(\lambda_0))\right\}$. Consequently, solving for $V_\beta(\lambda_0)$ we obtain

$$V_\beta(\lambda_0) = \max_{n\geq 0}\left\{\frac{R_1\frac{1-\beta^n}{1-\beta} + \beta^n(1-\tau)[(1-(1-\beta)g_n)R_1 + g_n R_2]}{1-\beta^{n+1}[1-(1-\beta)g_n]}\right\}. \tag{29}$$

If $\lambda_1 \geq \rho_2$ then $V_\beta(\lambda_1) = V_{\beta,T_h}(\lambda_1)$, *i.e.*, $V_\beta(\lambda_1) = \frac{\lambda_1 R_2 + \beta(1-\lambda_1)V_\beta(\lambda_0)}{1-\beta\lambda_1}$. And, using (25), $V_\beta(\lambda_0)$ is computed as follows $V_\beta(\lambda_0) = \max\{X_1, X_2\}$, where X_1 is given by (28) and X_2 is given by

$$X_2 = \max_{n\geq 0}\left\{\frac{[\frac{1-\beta^n}{1-\beta} + \beta^n(1-\tau)(1-\kappa_n)]R_1 + \beta^n[g_n - \tau\kappa_n]R_2}{1-\beta^{n+1}[1-(1-\beta)g_n]}\right\}. \tag{30}$$

Again, once $V_\beta(\lambda_0)$ and $V_\beta(\lambda_1)$ have been computed for the three scenarios, we retain the scenario that gives the maximal values.

C. Three thresholds policy

Since $\lambda_1 \geq \lambda_F \geq \rho_3$, we have $V_\beta(\lambda_1) = \frac{\lambda_1 R_2 + \beta(1-\lambda_1)V_\beta(\lambda_0)}{1-\beta\lambda_1}$ and $V_\beta(\lambda_0)$ is calculated as $V_\beta(\lambda_0) = \max\{X_1, X_2\}$, where X_1 is given by (28) and X_2 is given by (30).

Acknowledgement

The authors gratefully acknowledge the detailed comments and suggestions of Professor Qing Zhao (U.C. Davis).

References

[1] A. J. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1218-1230, Oct. 1997.

[2] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, pp. 1688-1692, Nov. 1999.

[3] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. Jou.*, vol. 39, pp. 1253-1265, Sept. 1960.

[4] L. Johnston and V. Krishnamurthy, "Opportunistic file transfer over a fading channel - A POMDP search theory formulation with optimal threshold policies," *IEEE Transactions Wireless Communications*, vol. 5, no. 2, pp. 394-405, Feb. 2006.

[5] D. Zhang and K. M. Wasserman, "Transmission schemes for time-varying wireless channels with partial state observations," *Proc. of INFOCOM*, pp. 467-476, 2002.

[6] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, vol. 55, no. 5, pp. 2294-2309, May, 2007.

[7] S. M. Ross, "Quality control under Markovian deterioration," *Management Science* vol. 17, no. 9, pp. 587-596, May 1971.

[8] E. L. Sernik and S. I. Marcus, "On the computation of the optimal cost function for discrete time Markov models with partial observations," *Annals of Operations Research*, vol. 29, pp. 471-512, Apr. 1991.

[9] G. E. Monahan "Optimal stopping in a partially observable binary-valued Markov chain with costly perfect information," *Journal of Applied Probability*, vol. 19, pp.72-81, 1982.

[10] R. Smallwood and E. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Ops. Research*, pp. 10711088, 1971.

[11] S. M. Ross, "Applied probability models with optimization applications," San Francisco: Holden-Day, 1970.

[12] A. Laourine and L. Tong, "Betting on Gilbert-Elliot Channels," Cornell University, Tech. Rep. ACSP TR-01-09-14, January 2009. [Online]. Available: http://acsp.ece.cornell.edu/papers/ACSP-TR-01-09-14.