Betting on Gilbert-Elliot Channels

Amine Laourine, *Student Member, IEEE*, and Lang Tong, *Fellow, IEEE*

***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 transmission with a high data rate but provides no protection against a bad channel, ii) betting conservatively by using a strong code that perfectly protects the communication against a bad channel but does not allow a high data rate, 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.**

***Index Terms*—Gilbert-Elliot channel, Opportunistic channel access, Markov decision processes.**

I. Introduction

COMMUNICATION over the wireless medium is subject to multiple impairments such as fading, path loss, and interference. These effects degrade the quality of service and lead to transmission failures. The quality of the radio channel is often random and evolves in time, ranging from good to bad depending on the propagation conditions. To cope with this changing behavior and maintain a good quality of service, multi-rate modulations or link adaptation may be performed. Link adaptation, also known as adaptive modulation and coding, is a technique that leads to a better channel utilization by matching the systems parameters of the transmitted signal (e.g., data/coding rate, constellation size and transmit power) to the changing channel conditions [1].

It is well established that 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 Markov 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 communication system operating over a Gilbert-Elliot channel in a time-slotted fashion. 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 using 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.

A. Related Work

MDP tools have been applied to solve communication problems over time-varying channels, see, *e.g.*, [4]-[7]. In [4], the authors considered rate and power control strategies for transmitting a fixed number of bits over fading channels subject to both energy and delay constraints. In [5], the authors obtained the optimal rate control policy in wireless networks with Rayleigh fading channels. Most related to this paper are [6] and [7]. In [6], 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 [7], 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 (or cognitive) spectrum access (refer to [9] for an overview) where sensing is an integral part of the access scheme. A generic setup is as follows: a cognitive (or secondary) user tries to opportunistically access a channel which, depending on the state of the primary user, can be either busy or idle. Relying on the theory of Partially Observable Markov Decision Processes (POMDP), several transmission and scheduling policies have been developed over the past years [8]-[14]. For instance, in [11], the authors derive optimal joint probing and transmission policies in multichannel wireless systems. In that work, however, the channel state is assumed to be independent from slot to slot. In [8], [10], [12]-[14], the authors target the problem of optimal access to multiple Gilbert-Elliot channels. In their setup, a sensing action is always carried out by the secondary user before attempting any transmission. The problem considered here is different in that the transmitter is allowed to transmit without

Manuscript received January 14, 2009; revised June 9, 2009 and October 18, 2009; accepted October 21, 2009. The associate editor coordinating the review of this paper and approving it for publication was G. Vitetta.

Part of this paper was presented at the IEEE Military Communications Conference (MILCOM) 2009.

The authors are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, USA (e-mail: {al496@, ltong}@ece.cornell.edu).

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

Digital Object Identifier 10.1109/TWC.2010.02.090055

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 [15], 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.

B. Main results and organization

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 [15]-[17] to establish the optimality 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 study the scenario in which the transmitter receives always the channel quality at the end of the slot, which corresponds to the situation when the receiver feeds back[1] the channel state information after each transmission. In Section VII, we provide numerical examples to illustrate the various theoretical results that will be presented in the paper. Finally, Section VIII 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 the channel transitions occur at the beginning of the time frame. We assume also 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 [6], [7]). 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.

[1]Note that this feedback channel provides extra information only in the case where the transmitter decides to use the strong code, since in the two other cases (sensing or using a weak code) the transmitter will know the channel conditions either from the sensing outcome or from the ACK/NAK received from the sink at the end of the slot. Refer to Section II.B for further details.

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. It is of course natural to assume that the transmission is successful if the channel is in the good state. Note 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. When there is no channel state information feedback from the receiver, this assumption means also that the transmitter will not acquire any knowledge about the channel state during the elapsed slot. The situation where the transmitter is informed of the channel state after selecting the action T_l will be treated separately in Section VI.

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 lasts a fraction $\tau(< 1)$ of the slot. We assume also that sensing is perfect, *i.e.*, sensing reveals the true state of the channel. 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. Finally, note that τ may have to be selected large enough so that the perfect sensing assumption holds.

Depending on the sensing outcome, the transmitter will send $(1-\tau)R_1$ 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 certain 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 [18], 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

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

where $\mathbb{E}_\pi$ represents the expectation given that the policy π is employed, 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 expected 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 $(1-\tau)[(1-X_t)R_1 + X_t R_2]$.

At first sight, it may seem that the expected discounted reward is inappropriate for our problem, since why would the transmitter have a preference for bits transmitted now over bits transmitted in the future. This formulation provides however a tractable solution, and one can gain insights into the optimal policy when β is close to 1. One can also view β as the probability that a particular user is allowed to use the channel (see [5] for further details). Finally, from Th.6.17 and Th.6.18 in [19], it can be seen also that the discounted reward criterion is of primary importance when it comes to the derivation for the optimal policy of the average reward criterion (throughput).

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 [19, 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 (the case with CSI feedback is treated in Section VI). 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 if the safe action is taken, the value function evolves as

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

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 if the sensing action is taken, the value function evolves as

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

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 if the risky action is taken, the value function evolves as

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

Finally the Bellman equation for our communication problem reads as follows

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

As a final remark, we refer the interested reader to [20] where we consider the impact of channel sensing errors. In Section VI, we will treat the case where the transmitter knows the channel state information (CSI) at the end of each slot through a feedback channel. Note that in the present section, the transmitter acquires this delayed CSI only if the S or

T_h actions are taken[2]. But, if the action T_l is taken instead, the CSI is not known since in all cases the transmission is successful.

III. Structure of the Optimal Policy

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

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

Proof: We first start by proving the convexity of the value function. Define $V_\beta(p,n)$ as the optimal value when the decision horizon spans only n stages. Then we have the following recursion

$$V_\beta(p,n) = \max\{V_{\beta,T_l}(p,n), V_{\beta,S}(p,n), V_{\beta,T_h}(p,n)\}, \quad (10)$$

with $V_{\beta,T_l}(p,n) = R_1 + \beta V_\beta(T(p), n-1)$, $V_{\beta,S}(p,n) = (1-\tau)[R_1 + p(R_2 - R_1)] + \beta[(1-p)V_\beta(\lambda_0, n-1) + pV_\beta(\lambda_1, n-1)]$ and $V_{\beta,T_h}(p,n) = pR_2 + \beta[(1-p)V_\beta(\lambda_0, n-1) + pV_\beta(\lambda_1, n-1)]$.

Note that $V_\beta(p,1) = \max\{R_1, (1-\tau)[R_1 + p(R_2 - R_1)], pR_2\}$ is a convex function since it is the maximum of three convex functions. Assume that $V_\beta(p, n-1)$ is convex, then for $a \in [0,1]$ we have

$$\begin{aligned} & R_1 + \beta V_\beta(T(ap_1 + (1-a)p_2), n-1) \\ = \;& R_1 + \beta V_\beta(aT(p_1) + (1-a)T(p_2), n-1) \\ \leq \;& R_1 + a\beta V_\beta(T(p_1), n-1) + (1-a)\beta V_\beta(T(p_2), n-1) \\ = \;& aV_{\beta,T_l}(p_1, n) + (1-a)V_{\beta,T_l}(p_2, n) \\ \leq \;& aV_\beta(p_1, n) + (1-a)V_\beta(p_2, n). \end{aligned} \quad (11)$$

Also since the second and third terms in (10) are linear, we can easily see that

$$V_\beta(ap_1 + (1-a)p_2, n) \leq aV_\beta(p_1, n) + (1-a)V_\beta(p_2, n). \quad (12)$$

Hence $V_\beta(p,n)$ is convex. And by induction we have convexity for all n. However, from the theory of MDPs, we know that $V_\beta(p,n) \to V_\beta(p)$ as $n \to \infty$. Hence $V_\beta(p)$ is convex.
The proof that $V_\beta(p)$ is nondecreasing is also done by induction. Indeed since $R_2 > R_1$, we have that $V_\beta(p,1)$ is the maximum of three nondecreasing functions and is hence nondecreasing. Assume that $V_\beta(p, n-1)$ is nondecreasing, since $\lambda_1 \geq \lambda_0$, we have $V_\beta(\lambda_0, n-1) \leq V_\beta(\lambda_1, n-1)$. Hence the second and the third terms in (10) are nondecreasing functions. Also since $T(p)$ is nondecreasing, we have $V_\beta(T(p), n-1)$ is nondecreasing. Thus $V_\beta(p,n)$ is the maximum of three nondecreasing functions and is hence nondecreasing. Consequently, by letting $n \to \infty$ we obtain the desired result. ∎
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 $0 \leq \rho_1 \leq \rho_2 \leq \rho_3 \leq 1$ such that*

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

[2]The CSI is acquired through the ACK/NAK received from the transmitter if the T_h action is chosen or through sensing if the S action is taken.

Proof: We introduce the following sets

$$\Phi_{\mathcal{K}} = \{p \in [0,1], V_\beta(p) = V_{\beta,\mathcal{K}}(p)\}, \quad \mathcal{K} \in \{T_l, T_h, S\}. \quad (13)$$

In other words, $\Phi_{\mathcal{K}}$ is the set of beliefs for which it is optimal to take the action $\mathcal{K}$. We will prove that Φ_{T_h} and Φ_S are convex, which implies the structure of the optimal policy. This proof parallels to that of Ross [15].

Let $p_1, p_2 \in \Phi_{T_h}$ and let $a \in [0,1]$ then we have

$$\begin{aligned} V_\beta(ap_1 + (1-a)p_2) &\leq aV_\beta(p_1) + (1-a)V_\beta(p_2) \\ &= aV_{\beta,T_h}(p_1) + (1-a)V_{\beta,T_h}(p_2) \\ &= V_{\beta,T_h}(ap_1 + (1-a)p_2) \\ &\leq V_\beta(ap_1 + (1-a)p_2), \end{aligned} \quad (14)$$

where the first inequality comes from the convexity of $V_\beta(p)$; the first equality follows from the fact that $p_1, p_2 \in \Phi_{T_h}$, and the last inequality from the definition of $V_\beta(\cdot)$. Consequently $V_\beta(ap_1 + (1-a)p_2) = V_{\beta,T_h}(ap_1 + (1-a)p_2)$, and hence $ap_1 + (1-a)p_2 \in \Phi_{T_h}$, which proves the convexity of Φ_{T_h}. Since convex subsets of the real line are intervals and $1 \in \Phi_{T_h}$, there exists $\rho_3 \in (0,1]$ such that $\Phi_{T_h} = [\rho_3, 1]$. Using the same technique we can prove that Φ_S is convex and hence there exits $\rho_1, \rho_2 \in [0,1]$ such that $\Phi_S = [\rho_1, \rho_2]$. Consequently we have also that $\Phi_{T_l} = [0, \rho_1) \cup (\rho_2, \rho_3)$. ∎

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 VII 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 exist 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. In the upcoming section we will make use of the following operators:

$$T^n(p) = T(T^{n-1}(p)) = \lambda_F(1 - \alpha^n) + \alpha^n p. \quad (15)$$

$$T^{-n}(p) = T^{-1}(T^{-(n-1)}(p)) = \frac{p}{\alpha^n} - \frac{1-\alpha^n}{1-\alpha}\frac{\lambda_0}{\alpha^n}. \quad (16)$$

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).
Before delving into the computation of the thresholds, we need

to introduce the following technical lemma.

Lemma 1. 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

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

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 give a complete characterization of the thresholds for each policy.

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

$$\rho = \frac{R_1}{R_2 + \beta V_\beta(\lambda_1) - \beta \frac{R_1}{1-\beta}}. \tag{17}$$

Otherwise, we have

$$\rho = \frac{(1-\beta\lambda_1)R_1 + \beta\lambda_0 R_2 + \beta(\beta-1)(1-\beta\alpha)V_\beta(\lambda_0)}{(1-\beta\alpha)(R_2 + \beta(\beta-1)V_\beta(\lambda_0))}. \tag{18}$$

Proof: The threshold ρ is the solution of the equation $R_1 + \beta V_\beta(T(\rho)) = V_{\beta,T_h}(\rho)$. We can distinguish two possible scenarios:
If[3] $\frac{R_1}{1-\beta} \geq V_{\beta,T_h}(\lambda_F)$, then we have $\lambda_F \leq \rho$ and $T(\rho) \leq \rho$. Consequently, from lemma 1 we deduce that $V_\beta(T(\rho)) = \frac{R_1}{1-\beta}$, hence solving for ρ we obtain (17). Otherwise, we have $\lambda_F > \rho$, consequently $T(\rho) > \rho$ and $V_\beta(T(\rho)) = V_{\beta,T_h}(T(\rho))$, hence solving for ρ we obtain (18). ∎

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 $\rho_2 = \frac{(1-\tau)R_1}{(1-\tau)R_1 + \tau R_2}$.

[3]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.

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 3. 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[4] $V_{\beta,T_l}(T^{-1}(\rho_2)) < V_{\beta,S}(T^{-1}(\rho_2))$, ρ_1 will be equal to (19) given at the bottom of this page.
Else ρ_1 will be equal to (20) 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 (19).
3) Finally if $\frac{R_1}{1-\beta} \geq V_{\beta,S}(\lambda_F)$ and $\lambda_F \leq \rho_2$, then

$$\rho_1 = \frac{\tau(1-\beta)R_1}{(1-\tau)(1-\beta)(R_2-R_1) + \beta((1-\beta)V_\beta(\lambda_1) - R_1)}. \tag{21}$$

Proof: The threshold ρ_1 is the solution to the equation $R_1 + \beta V_\beta(T(\rho_1)) = V_{\beta,S}(\rho_1)$. We can distinguish three possible scenarios:
1) If $\lambda_F > \rho_2$ then two cases can be distinguished: If $V_{\beta,T_l}(T^{-1}(\rho_2)) < V_{\beta,S}(T^{-1}(\rho_2))$ we will have $\rho_1 < T(\rho_1) \leq \rho_2$ and hence $V_\beta(T(\rho_1)) = V_{\beta,S}(T(\rho_1))$, consequently solving for ρ_1 we obtain (19). Else, $T(\rho_1) > \rho_2$ and consequently $V_\beta(T(\rho_1)) = V_{\beta,T_h}(T(\rho_1))$ and hence solving for ρ_1 we obtain (20).
2) If $\frac{R_1}{1-\beta} < V_{\beta,S}(\lambda_F)$ and $\lambda_F \leq \rho_2$, then it follows that $\rho_1 < \lambda_F \leq \rho_2$. Consequently, $\rho_1 < T(\rho_1) < \lambda_F$, *i.e.*, $V_\beta(T(\rho_1)) = V_{\beta,S}(T(\rho_1))$, and ρ_1 will be given by (19).
3) Finally if $\frac{R_1}{1-\beta} \geq V_{\beta,S}(\lambda_F)$ and $\lambda_F \leq \rho_2$, then we must have $\lambda_F \leq \rho_1$, *i.e.*, $T(\rho_1) < \rho_1$ and $V_\beta(T(\rho_1)) = \frac{R_1}{1-\beta}$ (c.f. lemma 1). Hence solving for ρ_1, we obtain (21). ∎

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 4. If a three-thresholds policy is optimal, then $\lambda_F \in [\rho_3, 1]$.

Proof: We first prove that $\lambda_F \notin [\rho_1, \rho_2]$. Note that both ρ_1 and ρ_2 satisfy the following equation

$$R_1 + \beta V_\beta(T(\rho)) = V_{\beta,S}(\rho). \tag{22}$$

So if $\lambda_F \in [\rho_1, \rho_2]$, then $T(\rho_1), T(\rho_2) \in [\rho_1, \rho_2]$, *i.e.*, $V_\beta(T(\rho_1)) = V_{\beta,S}(T(\rho_1))$, and the same for $V_\beta(T(\rho_2))$. Consequently, (22) would have a single solution given by (19), and we would have $\rho_1 = \rho_2$, this contradicts the assumption

[4]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 $V_{\beta,S}(T^{-1}(\rho_2))$.

$$\rho_1 = \frac{\tau R_1 + \beta(1-\tau)[R_1 + \lambda_0(R_2-R_1)] + \beta^2[V_\beta(\lambda_0) + \lambda_0(V_\beta(\lambda_1) - V_\beta(\lambda_0))] - \beta V_\beta(\lambda_0)}{(1-\beta\alpha)[(1-\tau)(R_2-R_1) + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0))]} \tag{19}$$

$$\rho_1 = \frac{\beta\lambda_0 R_2 + (1-\beta\lambda_1)\tau R_1 + \beta(\beta-1)(1-\beta\alpha)V_\beta(\lambda_0)}{R_2(1-\lambda_1(\beta(\alpha-\tau)-\tau)) + \beta(\beta-1)(1-\beta\alpha)V_\beta(\lambda_0) - (1-\tau)(1-\beta\lambda_1)R_1} \tag{20}$$

that $\rho_1 < \rho_2$.

Assume that $\lambda_F \in [0, \rho_1]$, then from lemma 1 we have that $V_\beta(p) = \frac{R_1}{1-\beta}$ for $p \in [0, \rho_1]$. Now for $p \in [\rho_2, \rho_3]$, we have $V_\beta(p) = R_1 + \beta V_\beta(T(p))$. However $T(p) \leq p$ for $p \geq \lambda_F$ and $V_\beta(\cdot)$ is increasing, hence $V_\beta(p) \leq R_1 + \beta V_\beta(p)$ or equivalently, $V_\beta(p) \leq \frac{R_1}{1-\beta}$. Remember that $V_\beta(p) \geq V_\beta(\lambda_F) = \frac{R_1}{1-\beta}$ (because $V_\beta(\cdot)$ is $\nearrow$). Consequently, $V_\beta(p) = \frac{R_1}{1-\beta}$ for $p \in [\rho_2, \rho_3]$, and for the same reasons for $p \in [\rho_1, \rho_2]$, *i.e.*, $V_\beta(p) = \frac{R_1}{1-\beta}$ for $p \in [0, \rho_3]$. This is a contradiction with the assumption that we have a three threshold policy. Finally, using the same reasoning we prove that $\lambda_F \notin [\rho_2, \rho_3]$. ■

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 (18). The two other thresholds ρ_1 and ρ_2 are computed as in the following lemma.

Lemma 5. If a 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

$$\begin{aligned}\gamma(k) &= [(1-\tau)(R_2 - R_1) + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0))]\frac{1}{\alpha^k} \\ &- \beta^k[R_2 + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0))],\end{aligned} \tag{23}$$

and

$$\begin{aligned}\delta(k) = R_1\frac{1-\beta^k}{1-\beta} + \beta(\beta^k - 1)V_\beta(\lambda_0) - (1-\tau)R_1 \\ + \frac{\lambda_0(1-\alpha^k)}{\alpha^k(1-\alpha)}[(1-\tau)(R_2 - R_1) + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0))].\end{aligned} \tag{24}$$

We have then that ρ_2 is equal to (25) 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 (19). Else, let $J'+1 = \min\{k \geq J+2 : \gamma(k)\rho_3 < \delta(k)\}$, and ρ_1 will be given by (25) with J replaced by J'.

Proof: Note first that there exists no $k \in \mathbb{N}$ such that $T^{-(k+1)}(\rho_3) < \rho_1 < \rho_2 < T^{-k}(\rho_3)$, for otherwise, we would have $V_{\beta,T_l}(\rho_1) = V_{\beta,T_l}(\rho_2)$ and (22) would have only one solution, thereby contradicting the three-thresholds assumption. Let $J+1 = \min\{k \in \mathbb{N} : T^{-k}(\rho_3) < \rho_2\}$, it follows then that $T^{-(J+1)}(\rho_3) < \rho_2 \leq T^{-J}(\rho_3)$, or equivalently

$$\begin{aligned}V_\beta(\rho_2) &= R_1 + \beta V_\beta(T(\rho_2)) = R_1 + \beta R_1 + \beta^2 V_\beta(T^2(\rho_2)) \\ &= \ldots = R_1\frac{1-\beta^{J+1}}{1-\beta} + \beta^{J+1}V_{\beta,T_h}(T^{J+1}(\rho_2)).\end{aligned} \tag{26}$$

Hence $V_\beta(\rho_2)$ will be given by

$$\begin{aligned}V_\beta(\rho_2) = R_1\tfrac{1-\beta^{J+1}}{1-\beta} + (\alpha\beta)^{J+1}(R_2 + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0)))\rho_2 \\ + \beta^{J+1}[P_F(1-\alpha^{J+1})(R_2 + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0))) + \beta V_\beta(\lambda_0)],\end{aligned} \tag{27}$$

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

Since ρ_2 is a solution to (22), we solve for ρ_2 to obtain (25). It is easily seen that $J+1 = \min\{k \in \mathbb{N} : V_{\beta,T_l}(T^{-k}(\rho_3)) < V_{\beta,S}(T^{-k}(\rho_3))\}$. The term $V_{\beta,T_l}(T^{-k}(\rho_3))$ can be calculated as follows

$$V_{\beta,T_l}(T^{-k}(\rho_3)) = R_1\frac{1-\beta^k}{1-\beta} + \beta^k[R_2\rho_3 + \beta[\rho_3 V_\beta(\lambda_1) + (1-\rho_3)V_\beta(\lambda_0)]]. \tag{28}$$

Similarly $V_{\beta,S}(T^{-k}(\rho_3))$ can be computed using (29) given below. Hence after some manipulations we obtain the expressions of $\gamma(k)$ and $\delta(k)$ as shown in the lemma.

Finally, if $V_{\beta,T_l}(T^{-1}(\rho_2)) < V_{\beta,S}(T^{-1}(\rho_2))$ we will have $\rho_1 < T(\rho_1) < \rho_2$, hence ρ_1 will be given by (19). Else we are in the situation where $T(\rho_1) > \rho_2$, hence by letting $J'+1 = \min\{k \geq J+2 : T^{-k}(\rho_3) < \rho_1\} = \min\{k \geq J+2 : V_{\beta,T_l}(T^{-k}(\rho_3)) > V_{\beta,S}(T^{-k}(\rho_3))\}$ and using the same approach used above we obtain the result presented in the lemma. ■

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 $p \in \Phi_{T_l}$.

A. One-threshold policy

The goal is to find $V_\beta(p)$ for $p \leq \rho$. Here we can distinguish two possibilities: If $\lambda_F \leq \rho$, then from lemma 1 we conclude that $V_\beta(p) = \frac{R_1}{1-\beta}$ for all $p \leq \rho$. 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 $1 \leq i \leq J$. Then for $p \in F_i$, we have $T^i(p) > \rho \geq T^{(i-1)}(p)$, *i.e.*,

$$\begin{aligned}V_\beta(p) &= R_1 + \beta V_\beta(T(p)) = R_1 + \beta R_1 + \beta^2 V_\beta(T^2(p)) \\ &= \ldots = R_1\frac{1-\beta^i}{1-\beta} + \beta^i V_{\beta,T_h}(T^i(p)).\end{aligned} \tag{30}$$

The optimal policy for this last case is illustrated in Fig. 1. As it can be seen the set of belief $[0, 1]$ is divided in two regions specified by the threshold ρ.

$$\rho_2 = \frac{R_1(\frac{1-\beta^{J+1}}{1-\beta} - (1-\tau)) + \beta^{J+1}[P_F(1-\alpha^{J+1})(R_2 + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0))) + \beta V_\beta(\lambda_0)] - \beta V_\beta(\lambda_0)}{(1-\tau)(R_2 - R_1) + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0)) - (\alpha\beta)^{J+1}(R_2 + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0)))} \tag{25}$$

$$V_{\beta,S}(T^{-k}(\rho_3)) = [(1-\tau)(R_2 - R_1) + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0))]\frac{\rho_3}{\alpha^k} + (1-\tau)R_1 + \beta V_\beta(\lambda_0) - \frac{\lambda_0(1-\alpha^k)}{\alpha^k(1-\alpha)}[(1-\tau)(R_2 - R_1) + \beta(V_\beta(\lambda_1) - V_\beta(\lambda_0))] \tag{29}$$

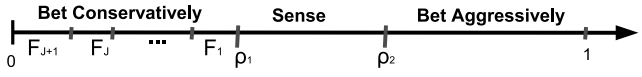

Fig. 2. Illustration of the two thresholds policy for $\rho_1 < \lambda_F \leq \rho_2$.

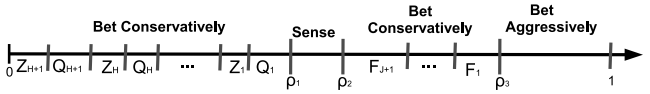

Fig. 3. Illustration of the three thresholds policy for $T(\rho_1) > \rho_2$ and $T^{-(H+1)}(\rho_2) \geq 0$.

B. Two-thresholds policy

The approach here is similar to the previous case, *i.e.*, if $\lambda_F \leq \rho_1$, then $V_\beta(p) = \frac{R_1}{1-\beta}$ for all $p \leq \rho_1$. If $\rho_1 < \lambda_F \leq \rho_2$, 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 $1 \leq i \leq J$. Then for $p \in F_i$, we have $\rho_2 > T^i(p) > \rho_1 \geq T^{(i-1)}(p)$, *i.e.*,

$$V_\beta(p) = R_1 \frac{1-\beta^i}{1-\beta} + \beta^i V_{\beta,S}(T^i(p)). \tag{31}$$

The optimal policy for this case is illustrated in Fig. 2, the interval $[0, 1]$ is separated in three regions by the thresholds ρ_1 and ρ_2.

If $\lambda_F > \rho_2$, two case can be distinguished: If $T(\rho_1) \leq \rho_2$ then the computation is similar to the situation where $\rho_1 < \lambda_F \leq \rho_2$ discussed above. If $T(\rho_1) > \rho_2$, let $F_{J+1} = [0, T^{-J}(\rho_1)]$, for $2 \leq i \leq J$ 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 $p \in F_i, i \geq 1$, $V_\beta(p)$ will be given by (31). For $p \in F_0 = (T^{-1}(\rho_2), \rho_1]$ we have

$$V_\beta(p) = R_1 + \beta V_{\beta,T_h}(T(p)). \tag{32}$$

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 $1 \leq i \leq J$. For $p \in F_i$, we have $T^i(p) \geq \rho_3$, *i.e.*, $V_\beta(p)$ is given by (30).

For $p \in [0, \rho_1]$ we can distinguish two cases: If $T(\rho_1) \leq \rho_2$, $V_\beta(p)$ for $p \in [0, \rho_1]$ is computed using (31). If $T(\rho_1) > \rho_2$, let $H+1 = \min\{k \in \mathbb{N} : T^{-k}(\rho_1) < 0\}$. Then we have two subcases: First, if $T^{-(H+1)}(\rho_2) \geq 0$, then let $Z_{H+1} = [0, T^{-(H+1)}(\rho_2))$, for $1 \leq i \leq H$ let $Z_i = [T^{-i}(\rho_1), T^{-i}(\rho_2))$ and for $1 \leq i \leq H+1$ 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 (31). For $p \in Q_i$, $T^i(p) \in [\rho_2, \rho_3)$, hence there exits $1 \leq j \leq J+1$ such that $T^i(p) \in F_j$, *i.e.*,

$$V_\beta(p) = R_1 \frac{1-\beta^{i+j}}{1-\beta} + \beta^{i+j} V_{\beta,T_h}(T^{i+j}(p)). \tag{33}$$

The optimal policy for this case is illustrated in Fig. 3. Note that when $\rho_1 = \rho_2$, the policy degenerates to the one threshold policy. Whereas if $\rho_2 = \rho_3$, the policy reduces to the two thresholds policy discussed above. Second, if $T^{-(H+1)}(\rho_2) < 0$, then let $Z_{H+1} = [0, T^{-H}(\rho_1))$, for $1 \leq i \leq H$ 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 (33). For $p \in Q_i, T^i(p) \in [\rho_1, \rho_2)$ and consequently $V_\beta(p)$ is computed using (31).

VI. Optimal Policy with Channel State Information feedback

A. Structure of the value function and of the optimal policy

In this section we consider the situation where the transmitter knows the channel state information (CSI) at the end of each slot. Note that in the previous model, the transmitter acquires this delayed CSI only if the S or T_h actions are taken. But, if the action T_l is taken instead, the CSI is not known since in all cases the transmission is successful. Now, in this new model, we assume that if the action T_l is taken, the receiver replies back with the CSI. This CSI feedback can take the form of one bit; 0 indicating a bad channel and 1 for a good channel. If the transmitter receives a 0, then the correct action (i.e., T_l) has been selected. Whereas if a 1 is received, the transmitter is informed that an opportunity of sending more data has been missed (if the action T_h was selected instead of T_l).

In this new model, $V_{\beta,T_l}(p)$ changes to

$$V_{\beta,T_l}(p) = R_1 + \beta((1-p)V_\beta(\lambda_0) + pV_\beta(\lambda_1)), \tag{34}$$

whereas $V_{\beta,S}(p)$ and $V_{\beta,T_h}(p)$ rest unchanged and as usual $V_\beta(p) = \max\{V_{\beta,T_l}(p), V_{\beta,S}(p), V_{\beta,T_h}(p)\}$. Recall in the previous model that $V_\beta(\cdot)$ is convex. Hence $V_\beta(T(p)) \leq (1-p)V_\beta(\lambda_0) + pV_\beta(\lambda_1)$, which proves that $V_\beta(\cdot)$ with CSI feedback is bigger than $V_\beta(\cdot)$ with no CSI feedback.

The optimal policy is easily obtained and is given in the following theorem (the proof is omitted).

Theorem 3. *If* $\frac{\tau R_1}{(1-\tau)(R_2-R_1)} > \frac{(1-\tau)R_1}{(1-\tau)R_1+\tau R_2}$ *then the optimal policy is a one threshold policy, i.e.,*

$$\pi^*(p) = \begin{cases} T_l & \text{if } 0 \leq p \leq \frac{R_1}{R_2} \\ T_h & \text{if } \frac{R_1}{R_2} \leq p \leq 1 \end{cases}.$$

If $\frac{\tau R_1}{(1-\tau)(R_2-R_1)} \leq \frac{(1-\tau)R_1}{(1-\tau)R_1+\tau R_2}$ *then the optimal policy is a two thresholds policy, i.e.,*

$$\pi^*(p) = \begin{cases} T_l & \text{if } 0 \leq p \leq \frac{\tau R_1}{(1-\tau)(R_2-R_1)} \\ S & \text{if } \frac{\tau R_1}{(1-\tau)(R_2-R_1)} \leq p \leq \frac{(1-\tau)R_1}{(1-\tau)R_1+\tau R_2} \\ T_h & \text{if } \frac{(1-\tau)R_1}{(1-\tau)R_1+\tau R_2} \leq p \leq 1 \end{cases}.$$

As one should expect, the optimal strategy here is a myopic policy, i.e., a policy that maximizes the immediate reward. Indeed, the optimal policy for this problem is identical to the optimal policy corresponding to the following MDP: $W_\beta(p) = \max\{R_1, (1-\tau)[(1-p)R_1 + pR_2], pR_2\}$.

B. Value function

Note that the value function is totally determined by finding $V_\beta(\lambda_0)$ and $V_\beta(\lambda_1)$. In order to determine the optimal action when the belief is λ_1 or λ_0, we compare these values to

the thresholds established above. Then all that remains is solving a system of two linear equations with two unknowns (i.e., $V_\beta(\lambda_0)$ and $V_\beta(\lambda_1)$). To illustrate the procedure of determining $V_\beta(\lambda_0)$ and $V_\beta(\lambda_1)$, we consider here the example where the optimal policy is a one threshold policy and $\lambda_0 \leq \frac{R_1}{R_2} \leq \lambda_1$. We have then

$$V_\beta(\lambda_0) = R_1 + \beta(V_\beta(\lambda_0) + (V_\beta(\lambda_1) - V_\beta(\lambda_0))\lambda_0), \quad (35)$$

$$V_\beta(\lambda_1) = \lambda_1 R_2 + \beta(V_\beta(\lambda_0) + (V_\beta(\lambda_1) - V_\beta(\lambda_0))\lambda_1). \quad (36)$$

Solving for $V_\beta(\lambda_0)$ and $V_\beta(\lambda_1)$ leads to

$$V_\beta(\lambda_0) = \frac{(1-\beta\lambda_1)R_1 + \beta\lambda_0\lambda_1 R_2}{(1-\beta)(1-\beta\alpha)}, \quad (37)$$

$$V_\beta(\lambda_1) = \frac{\beta(1-\lambda_1)R_1 + (1-\beta+\beta\lambda_0)\lambda_1 R_2}{(1-\beta)(1-\beta\alpha)}. \quad (38)$$

All other cases are treated similarly, and the details are omitted due to space limitations.

VII. Numerical results

We start by analyzing the scenario with no CSI feedback. We will consider three different setups, each 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, 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 $\tau = 0$.

Fig. 8 shows the impact of the sensing action on the overall performance. The system parameters in this plot are as follows: $R_1 = 3$, $R_2 = 4$, $\tau = 0.1$, $\beta = 0.9$, $\lambda_0 = 0.6$ and $\lambda_1 = 0.9$. We consider here three different scenarios: (i) the transmitter does not know the CSI at the end of each slot;[5], (ii) the transmitter has access to the CSI at the end of each slot (see Section VI), and (iii) the transmitter has access to the delayed CSI but can only use the actions T_l and T_h. As it can be seen in this example, the total number of transmitted bits is reduced in the third scenario. However, when the transmitter can access the sensing action, the total number of transmitted bits is substantially augmented and the optimal policy performs closely to the case with full CSI feedback.

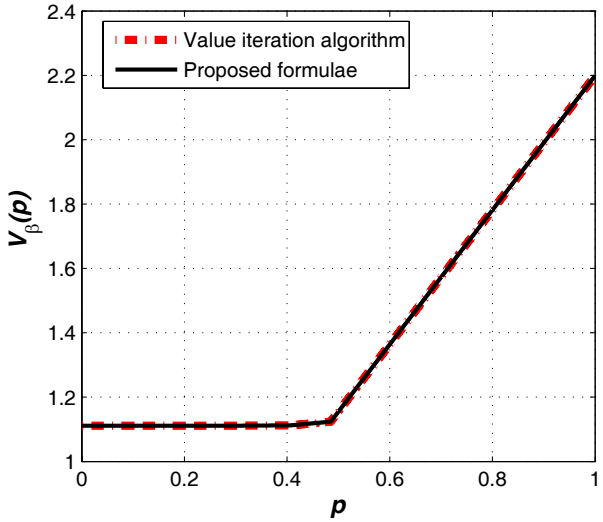

Fig. 4. Optimality of a one threshold policy.

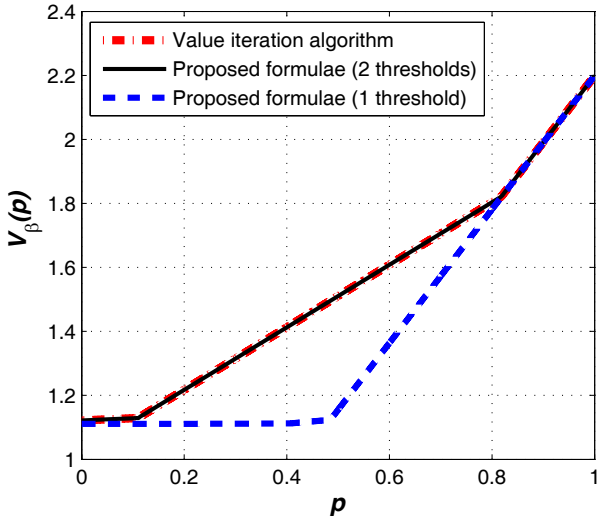

Fig. 5. Optimality of a two thresholds policy.

[5]Recall that in this paper, the transmitter always acquires this delayed CSI if the S or T_h actions are taken. But, if the action T_l is taken instead, the CSI is known only if there is a feedback channel from the receiver to the transmitter (see Section VI).

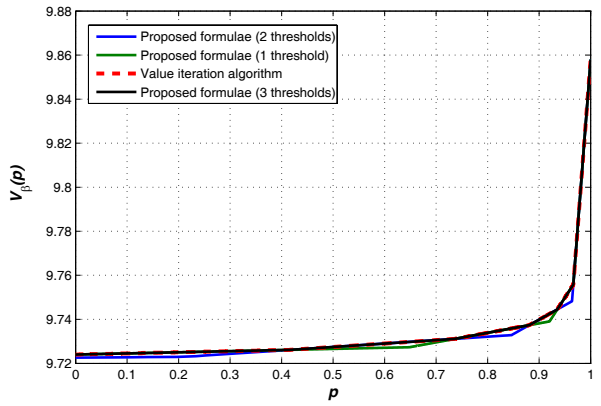

Fig. 6. Optimality of a three thresholds policy.

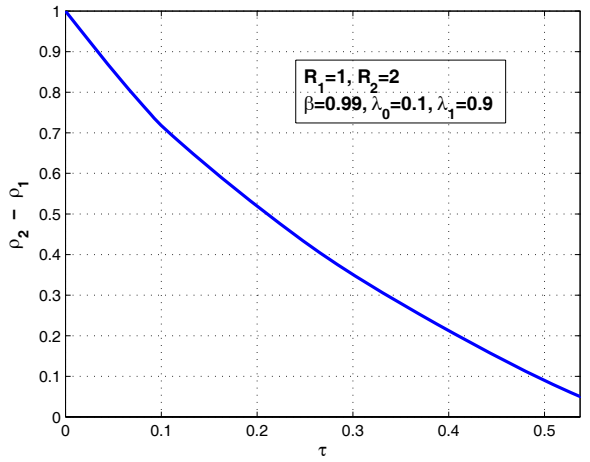

Fig. 7. The effect of the sensing duration τ on Φ_S.

VIII. 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 with a high data rate with no protection against a bad channel, ii) transmit with a low data rate 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 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) [21].

We have left some interesting problems open. We have not considered the design of the optimal policy when the communication system can communicate with more than two different rates. Another possibility is the extension of the problem to a multiple channel setup.

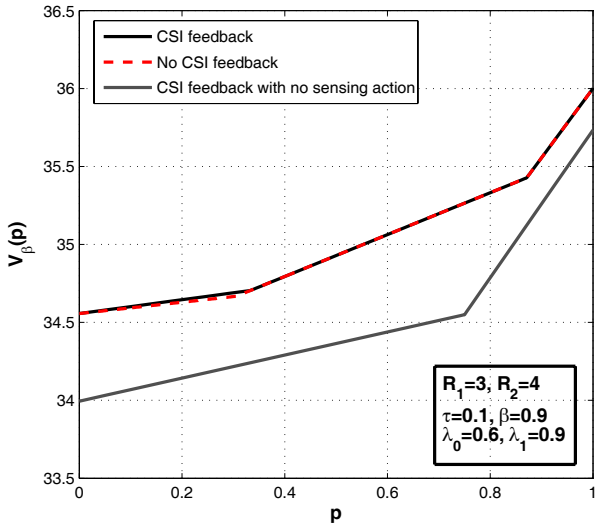

Fig. 8. Value function with and without CSI feedback.

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 4. *The value function can be written as*

$$V_\beta(p)=\max_{n\geq 0}\left\{\frac{1-\beta^n}{1-\beta}R_1+\beta^n \max\{V_{\beta,S}(T^n(p)),V_{\beta,T_h}(T^n(p))\}\right\}. \tag{39}$$

Proof: Recall that we have

$$V_\beta(p)=\max\{R_1+\beta V_\beta(T(p)),V_{\beta,S}(p),V_{\beta,T_h}(p)\}. \tag{40}$$

By replacing $V_\beta(T(p))$ by its expression we obtain

$$\begin{aligned}V_\beta(p)=\max\{&R_1(1+\beta)+\beta^2V_\beta(T^2(p)),R_1+\beta V_{\beta,S}(T(p)),\\&R_1+\beta V_{\beta,T_h}(T(p)),V_{\beta,S}(p),V_{\beta,T_h}(p)\}.\end{aligned} \tag{41}$$

Iterating over the same steps, we have for all $N\geq 0$ that

$$\begin{aligned}V_\beta(p)=\max\{&R_1\frac{1-\beta^N}{1-\beta}+\beta^NV_\beta(T^N(p)),\max_{0\leq n\leq N-1}\{\frac{1-\beta^n}{1-\beta}R_1\\&+\beta^n\max\{V_{\beta,S}(T^n(p)),V_{\beta,T_h}(T^n(p))\}\}\}.\end{aligned} \tag{42}$$

Since N is arbitrary and $0\leq\beta<1$, letting $N\to\infty$ we obtain the desired result. ■

Intuitively the previous result can be explained as follows; The expression $\frac{1-\beta^n}{1-\beta}R_1+\beta^nV_{\beta,S}(T^n(p))$ 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.

We now proceed with the computation of $V_\beta(\lambda_1)$ and $V_\beta(\lambda_0)$ by considering each policy separately.

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 1, 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 (39), we have that $V_\beta(\lambda_0)$ is a solution to the following equation

$$\begin{aligned} V_\beta(\lambda_0) &= \max_{n\geq 0}\left\{\frac{1-\beta^n}{1-\beta}R_1 + \beta^n V_{\beta,T_h}(T^n(\lambda_0))\right\} \\ &= \max_{n\geq 0}\{\frac{1-\beta^n}{1-\beta}R_1 + \beta^n(\kappa_n R_2 + \beta(V_\beta(\lambda_0) \\ &+ \kappa_n(V_\beta(\lambda_1) - V_\beta(\lambda_0))))\}, \end{aligned} \tag{43}$$

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

$$V_\beta(\lambda_0) = \max_{n\geq 0}\left\{\frac{\frac{1-\beta^n}{1-\beta}R_1 + \beta^n g_n R_2}{1-\beta^{n+1}[1-(1-\beta)g_n]}\right\}, \tag{44}$$

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: First, if $\lambda_1 \leq \rho_1$, then from lemma 1 we deduce that $V_\beta(\lambda_1) = V_\beta(\lambda_0) = \frac{R_1}{1-\beta}$. Second, 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 (39) 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

$$V_\beta(\lambda_0) = \max_{n\geq 0}\left\{\frac{R_1\frac{1-\beta^n}{1-\beta} + \beta^n(1-\tau)[(1-(1-\beta)g_n)R_1 + g_n R_2]}{1-\beta^{n+1}[1-(1-\beta)g_n]}\right\}. \tag{45}$$

Finally, 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 (39), $V_\beta(\lambda_0)$ is computed as follows $V_\beta(\lambda_0) = \max\{X_1, X_2\}$, where X_1 is given by (44) and X_2 is given by

$$X_2 = \max_{n\geq 0}\left\{\frac{[\frac{1-\beta^n}{1-\beta} + \beta^n(1-\tau)(1-\kappa_n)]R_1 + \beta^n[g_n - \tau\kappa_n]R_2}{1-\beta^{n+1}[1-(1-\beta)g_n]}\right\}. \tag{46}$$

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

If the three thresholds policy is optimal, then from lemma 4, we know that $\lambda_F \geq \rho_3$. Hence, since $\lambda_1 \geq \lambda_F$, we conclude that $V_\beta(\lambda_1) = V_{\beta,T_h}(\lambda_1)$ which implies that $V_\beta(\lambda_1) = \frac{\lambda_1 R_2 + \beta(1-\lambda_1)V_\beta(\lambda_0)}{1-\beta\lambda_1}$. Finally, $V_\beta(\lambda_0)$ is calculated as $V_\beta(\lambda_0) = \max\{X_1, X_2\}$, where X_1 is given by (44) and X_2 is given by (46).

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. J.*, vol. 39, pp. 1253-1265, Sep. 1960.
[4] H. Wang and N. B. Mandayam, "Opportunistic file transfer over a fading channel under energy and delay constraints," *IEEE Trans. Commun.*, vol. 53, no. 4, pp. 632-644, Apr. 2005.
[5] J. Razavilar, K. J. R. Liu, and S. I. Marcus, "Jointly optimized bit-rate/delay control policy for wireless packet networks with fading channels," *IEEE Trans. Commun.*, vol. 50, no.3, pp. 484-494, Mar. 2002.
[6] L. Johnston and V. Krishnamurthy, "Opportunistic file transfer over a fading channel—a POMDP search theory formulation with optimal threshold policies," *IEEE Trans. Wireless Commun.*, vol. 5, no. 2, pp. 394-405, Feb. 2006.
[7] D. Zhang and K. M. Wasserman, "Transmission schemes for time-varying wireless channels with partial state observations," in *Proc. INFOCOM*, pp. 467-476, 2002.
[8] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework," *IEEE J. Sel. Areas Commun.*, special issue on adaptive, spectrum agile and cognitive wireless networks, vol. 25, no. 3, pp. 589-600, Apr. 2007.
[9] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 55, no. 5, pp. 2294-2309, May 2007.
[10] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2053-2071, May 2008.
[11] N. B. Chang and M. Liu, "Optimal channel probing and transmission scheduling for opportunistic spectrum access," in *Proc. MOBICOM*, pp. 27-38, 2007.
[12] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multi-channel opportunistic access: structure, optimality, and performance," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5431-5440, Dec. 2008.
[13] 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.
[14] S. H. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of myopic sensing in multi-channel opportunistic access," to appear in *IEEE Trans. Inf. Theory*.
[15] S. M. Ross, "Quality control under Markovian deterioration," *Management Science*, vol. 17, no. 9, pp. 587-596, May 1971.
[16] 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.
[17] G. E. Monahan, "Optimal stopping in a partially observable binary-valued Markov chain with costly perfect information," *J. Applied Probability*, vol. 19, pp. 72-81, 1982.
[18] R. Smallwood and E. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Ops. Research*, pp. 10711088, 1971.
[19] S. M. Ross, *Applied Probability Models with Optimization Applications*. San Francisco: Holden-Day, 1970.
[20] A. Laourine and L. Tong, "Betting on Gilbert-Elliot channels," Cornell University, Tech. Rep. ACSP TR-01-09-14, Jan. 2009. [Online]. Available: http://acsp.ece.cornell.edu/papers/ACSP-TR-01-09-14.
[21] H. Holma and A. Toskala, *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*. John Wiley & Sons, 3rd ed., 2004.

Amine Laourine (S'07) received the Diplome d'ingenieur from the Ecole Polytechnique de Tunisie (Tunisia Polytechnic School) in 2005 and a Master degree in telecommunications from the Institut National de la Recherche Scientifique (INRS) in 2007. Since August 2007 he has been a Ph.D. student in Electrical Engineering at Cornell University.

He received the best paper award at the wireless communication symposium in Globecom'07, the Irwin and Joan Jacobs's fellowship from Cornell University in 2007, the Rene-Fortier scholarship from Bell Canada and the Tunisian Government fellowship for academic excellence in 2006. Amine Laourine's main research interests are in the field of wireless communications and information theory.

Lang Tong (S'87,M'91,SM'01,F'05) is the Irwin and Joan Jacobs Professor in Engineering at Cornell University Ithaca, New York. He received the B.E. degree from Tsinghua University, Beijing, China, in 1985, and M.S. and Ph.D. degrees in electrical engineering in 1987 and 1991, respectively, from the University of Notre Dame, Notre Dame, Indiana. He was a Postdoctoral Research Affiliate at the Information Systems Laboratory, Stanford University in 1991. He was the 2001 Cor Wit Visiting Professor at the Delft University of Technology and had held visiting positions at Stanford University, and U.C. Berkeley.

Lang Tong is a Fellow of IEEE. He received the 1993 Outstanding Young Author Award from the IEEE Circuits and Systems Society, the 2004 best paper award (with Min Dong) from IEEE Signal Processing Society, and the 2004 Leonard G. Abraham Prize Paper Award from the IEEE Communications Society (with Parvathinathan Venkitasubramaniam and Srihari Adireddy). He is also a coauthor of six student paper awards. He received Young Investigator Award from the Office of Naval Research.

Lang Tong's research is in the general area of statistical signal processing, wireless communications and networking, and information theory. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON INFORMATION THEORY, and IEEE SIGNAL PROCESSING LETTERS.