Coding Rate Adaptation for Hybrid ARQ Systems over Time Varying Fading Channels with Partially Observable State

Ashok K. Karmokar, and Vijay K. Bhargava

Department of Electrical and Computer Engineering, 2356 Main Mall,
University of British Columbia, Vancouver, BC, V6T 1Z4, Canada.
E-mail: {ashokk, vijayb}@ece.ubc.ca

***Abstract*— We present a cross-layer optimization problem where the coding rate of hybrid ARQ systems is adapted with channel, buffer, and input traffic state to minimize packet errors as well as buffer delay. Representing both the incoming traffic and the time-varying wireless channel as a finite state Markov chain, it is shown that the problem forms a partially observable Markov decision process (POMDP) problem. Since finding optimal policy is PSPACE complete, we investigate two policy-heuristic approaches for the purpose of efficient and real time solving of our formulated POMDP problem. Numerical results reveal that the performances of these two heuristic are almost same as the case when channel states are fully observable.**

I. INTRODUCTION

Future wireless networks are envisioned to support high data rates, packet oriented transport, multimedia traffic and wide range of quality of services (QoS), e.g., throughput, delay, bit error rate (BER), etc requirements. One of the major problems of wireless environment is randomly fluctuation of the channel gain due to various unpredictable phenomena, such as mobility of the terminal, fading, shadowing, etc and errors occur in bursts. However, the quality of the channel at a particular block depends on the previous channel conditions due to significant degree of correlation of some of these impairments. Hence, the memory of the wireless channel is important to take into consideration [1]. Adaptive hybrid automatic repeat request (HARQ) scheme is a promising technique for increasing both the throughput and the reliability of the packet transmission over time-varying fading channel [2]. HARQ schemes include parity bits for both error detection and forward error correction (FEC). Whereas real-time traffic (e.g., voice and video) have stringent delay requirement, non real-time traffic, such as data can tolerate substantial delay. In the recent literatures (e.g., [3]-[6] and the references therein), there has been a considerable interest in adaptive HARQ schemes due to increased delay tolerance of many applications, such as file transfer, web browsing, messaging, etc. Most of the authors in those references propose to change the operation mode and correspondingly adapt the code rate based only on the channel state information in some ad hoc basis. In these papers, higher layer buffer delays have not been taken into consideration. But from the practical point of view, queueing delay and packet dropping should be considered for retransmission systems, where packets are stored in a finite sized buffer before transmission. Recently, it has been realized that the system performance can be improved significantly by adapting transmission parameters, not only with channel conditions but also with higher layer parameters.

We consider a cross-layer optimization problem of HARQ systems, where physical layer FEC coding is combined with data link layer automatic repeat request (ARQ), and is adapted with time-varying fading channel condition, buffer occupancy and input traffic to minimize both the buffer delay and the packet errors. Since these two objectives are conflicting, we consider to minimize a weighted sum and formulate the problem as a partially observable Markov decision process (POMDP) problem. Note that minimizing buffer delay means that scheduler has to send more data packets in a codeword, hence maximizing throughput. We assume that the present channel state is hidden at the transmitter, but it can be observed partially through the observation feedback of the decoded packets from the receiver. This makes the system state partially observable. In the present context, feedback positive acknowledgment (ACK) and negative acknowledgment (NAK) are the observations. The underlying fading channel is modeled as correlated finite state Markov channel (FSMC). Unlike above mentioned references, the proposed scheme doesn't require channel estimation by counting the number of NAKs. Instead, control actions are chosen based on the *belief* of the states that can be tracked from the observations and previously taken actions. Since finding optimal solution of the POMDP problem is extremely computationally demanding, we explore two policy-heuristic to find the approximate solution of the POMDP problem.

The paper is organized as follows. Section II describes system model and channel model adopted in the paper. Formulation of the problem as a POMDP and its solution techniques are discussed in Section III. Section IV provides numerical results and discussions. We conclude in Section V.

II. MODEL DESCRIPTION

A. System Model

We consider a single mobile terminal communicating over a time-varying fading channel with partially observable state as shown in Fig. 1. Assume that the packets are coming from a higher layer application and are stored into the finite buffer for transmission in the subsequent time-slots. At a particular time-slot n, depending on the believed channel condition and current buffer occupancy, the controller decides the number of

0-7803-8938-7/05/$20.00 (C) 2005 IEEE

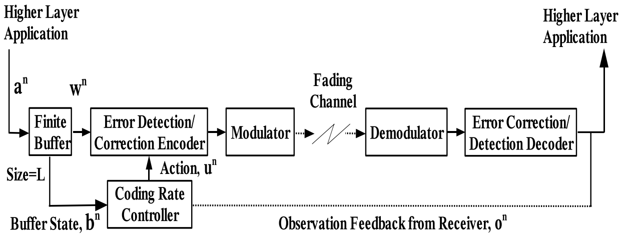

Fig. 1. Schematic of the Adaptive Type-I HARQ Systems

packets, w^n to be transmitted and encodes the corresponding $k = w^n E$ bits using a high rate (k, k') error detection code, where k is the number of information bits in the data packets of size E. These encoded packets are subsequently encoded using outer FEC code of rate k'/n' into codeword, where n' is the total number of bits in the final codeword. The resulting codeword is modulated and transmitted over the fading channel. After being received at the receiver, the block is demodulated. The decoder first attempts to correct any errors in the received demodulated block, then the decoded block is checked for error detection. If no errors are detected, the packets are delivered to the higher layer and an ACK is sent to the transmitter. Otherwise, the receiver discards the packets and sends a NAK requesting a retransmission of the same packets. The process is repeated until the packet is successfully received. We assume that the feedback channel is noiseless.

Let the incoming traffic between time-slot $n-1$ and n be a^n packets. It is assumed that packet arrival is independent of the channel fading and noise processes. We model the incoming traffic as a first order ergodic Markov chain. Let $\mathcal{A} = \{a_0, a_1, \ldots, a_{J-1}\}$ be the state space of incoming traffic with transition probability $p^a(a_j|a_i)$ from state a_i to state a_j, where a_i corresponds to i packets arrival. In a special case, if the packet arrival probability in a time-slot is q and is independent of previous packet arrivals, it is Bernoulli distributed. Hence, $\{a^n\}$ can be represented as a simple two state Markov chain, where the state space is $\mathcal{A} = \{a_0, a_1\}$. State a_0 and a_1 correspond to 0 and A packets arrival respectively.

Let $\mathcal{B} = \{b_0, b_1, \cdots, b_L\}$ denotes the state space of the buffer's packet occupancy, where b_i corresponds to i packets in the buffer. Let b^n denotes the number of packets in the buffer at the start of the nth block. The dynamics of the buffer is then given by,

$$b^{n+1} = \min\left\{b^n - \alpha w^n + a^{n+1}, b_L\right\}, \tag{1}$$

where the constraint on w^n is $0 \le w^n \le b^n$. The value of factor α equals 1 if ACK is received and 0 if NAK is received.

B. Channel Model

We adopt a FSMC model of the flat-fading Rayleigh channel, where the received power gain is partitioned into finite C number of non-overlapping states. First order Markov model is accurate enough for block/packet level communication when the block length is large [7] and used to capture the memory of the burst-error flat-fading channel for packet level communication (e.g., in [1]). Suppose the instantaneous received power gain of the channel be γ, then the probability density function of γ can be written as,

$$p(\gamma) = \frac{1}{\bar{\gamma}} e^{-\frac{\gamma}{\bar{\gamma}}} \text{ for } \gamma \ge 0, \tag{2}$$

where $\bar{\gamma} = \mathbb{E}\{\gamma\}$ is the average power gain of the channel. Let $\mathcal{C} = \{c_1, \cdots, c_C\}$ denote the state space of the channel and $\Gamma = \{\gamma_0, \cdots, \gamma_C\}$ is the corresponding received power gain thresholds in increasing order with $\gamma_0 = 0$ and $\gamma_C = \infty$. Then the channel is said to be in state $c_i, i = 1, \cdots, C$ if the received gain is between γ_{i-1} and γ_i. The average gain for state c_i is found as,

$$\tilde{\gamma}_i = \frac{\int_{\gamma_{i-1}}^{\gamma_i} \gamma p(\gamma) d\gamma}{\int_{\gamma_{i-1}}^{\gamma_i} p(\gamma) d\gamma} = \bar{\gamma} + \frac{\gamma_{i-1} e^{(\gamma_i - \gamma_{i-1})/\bar{\gamma}} - \gamma_i}{e^{(\gamma_i - \gamma_{i-1})/\bar{\gamma}} - 1}. \tag{3}$$

Now, the average duration of a state $c_i \in \mathcal{C}$ is expressed as,

$$\bar{\tau}_i = \frac{\text{Prob}\{\gamma_{i-1} \le \gamma \le \gamma_i\}}{N_{i-1} + N_i} = \frac{\pi_i}{N_{i-1} + N_i}, \tag{4}$$

where, N_i is the level crossing rate at the corresponding gain threshold γ_i and is given by,

$$N_i = \sqrt{\frac{2\pi\gamma_i}{\bar{\gamma}}} f_m e^{-\gamma_i/\bar{\gamma}}, \ \forall i, \tag{5}$$

and π_i is the steady-state probability of state c_i. It can be defined as,

$$\pi_i = \int_{\gamma_{i-1}}^{\gamma_i} p(\gamma) d\gamma = e^{-\frac{\gamma_{i-1}}{\bar{\gamma}}} - e^{-\frac{\gamma_i}{\bar{\gamma}}}. \tag{6}$$

In (5), $f_m = v/\lambda$ is the maximum Doppler frequency, where v and λ denote the speed of the mobile terminal and the wavelength of the radio wave respectively. In this paper, we use equal duration method of [8] for partitioning the gains. This method was shown to give better representation of the fading channel than other methods [8]. Therefore, we can write,

$$\bar{\tau}_i = r T_B, \text{ for } i = 1, \cdots, C, \tag{7}$$

where T_B is the block time period and r is a constant that should be larger than 1. Combining (4)-(7), we obtain

$$(r_1\sqrt{\gamma_{i-1}} - 1)e^{-\gamma_{i-1}/\bar{\gamma}} + (r_1\sqrt{\gamma_i} + 1)e^{-\gamma_i/\bar{\gamma}} = 0, \tag{8}$$

where $i = 1, \cdots, C$ and $r_1 = \sqrt{\frac{2\pi}{\gamma_0}} r f_m T_B$. Now, for a given value of C, f_m and T_B, the set of C nonlinear equations found from (8) can be solved numerically to get the value of r and $\gamma_1, \cdots, \gamma_{C-1}$. Let $p^c(c_j|c_i)$ denotes the transition probability from channel state c_i to state c_j. Now, if the fading rate of the channel is slow, it can be assumed that the quantized fading gain can jump only to the adjacent state c_j from state c_i, and the transition probabilities can be written as,

$$p^c(c_{i+1}|c_i) \approx \frac{N_i T_B}{\pi_i}, \ i = 1, \cdots, C-1 \quad \text{and} \tag{9}$$

$$p^c(c_{i-1}|c_i) \approx \frac{N_{i-1} T_B}{\pi_i}, \ i = 2, \cdots, C \tag{10}$$

The probability of staying in the same channel state can be found from the following equation,

$$p^c(c_i|c_i) = 1 - \sum_{j=i-1, j\neq i}^{i+1} p^c(c_j|c_i); i = 1, 2, \cdots, C. \quad (11)$$

In (11), redundant probabilities $p^c(c_0|c_1) = p^c(c_{C+1}|c_C) = 0$.

III. POMDP Formulation and Solution Techniques

A POMDP problem has the following elements: a set of time-slots $\mathcal{T} = \{1, 2, \cdots, m\}$, a set of system states $\mathcal{S}$, a set of actions $\mathcal{U}$, a set of transition probabilities $\mathcal{P}$, a set of observations $\mathcal{O}$, a set of observation probabilities Ω, and a set of costs $\mathcal{G}$. Let $\mathcal{S} = \mathcal{B} \times \mathcal{C} \times \mathcal{A} = \{s_1, s_2, \cdots, s_N\}$ denotes the composite state space for our model. The action space $\mathcal{U}_{s_i}$ is the set of possible choices of coding rates in state s_i. Each action $u_i \in \mathcal{U} = \{u_1, u_2, \cdots, u_U\}$ corresponds one-to-one a FEC code (n', k') of rate k'/n', where U denotes the number of actions and u_i denotes transmission of $i-1$ packets. Let $\mathcal{O} = \{o_1, o_2\}$ be the set of observations, where o_1 corresponds to the ACK and o_2 corresponds to NAK. Therefore the transition probabilities, costs and observation probabilities can be expressed as: $\mathcal{P} : \mathcal{S} \times \mathcal{U} \mapsto \Pi(\mathcal{S})$, $\mathcal{G} : \mathcal{S} \times \mathcal{U} \mapsto \mathbb{R}$ and $\Omega : \mathcal{S} \times \mathcal{U} \mapsto \Pi(\mathcal{O})$, where $\Pi(\cdot)$ represents the set of discrete probability distributions over a finite set. The immediate cost at time-slot n for the problem at hand can be given as the weighted sum of two objectives,

$$g(s^n, u^n) = \beta p(o_2|c^n, u^n) w^n + b^n/q, \quad (12)$$

where weighting factor $\beta \geq 0$, and $p(o_2|c^n, u^n)$ is the NAK probability, for a given channel state c^n and FEC (n', k', t) capable of correcting t bit errors, it can be written as,

$$p(o_2|c^n, u^n) = \sum_{l=t+1}^{n'} \binom{n'}{l} (P_{e,i})^l (1 - P_{e,i})^{n'-l}, \quad (13)$$

where, $P_{e,i}$ is the bit-error probability for channel state c_i. For MQAM modulation, valid for both low and high SNR, it can be expressed approximately as [9],

$$P_{e,i} = 4\left(\frac{2^{\frac{v}{2}} - 1}{v 2^{\frac{v}{2}}}\right) \sum_{i=1}^{2^{\frac{v}{2}-1}} Q\left((2i-1)\sqrt{\frac{3\tilde{\gamma} P_t v}{\sigma^2 (2^v - 1)}}\right), \quad (14)$$

where P_t is the transmitted power, σ^2 is the noise variance of the channel, and v is the number of bits in a symbol. Here, we assume that error detection code can detect all the remaining errors. Note that $p(o_2|s^n, u^n) = p(o_2|c^n, u^n)$ for all buffer states. The state transition probability for all $u_i \in \mathcal{U}_{s_i}$ and $s_i, s_j \in \mathcal{S}$ can be written as,

$$\begin{aligned} p_{s_i,s_j}(u_i) &= p^c(\chi(s_j)|\chi(s_i)) \sum_{o_i \in \mathcal{O}} \sum_{a_i \in \mathcal{A}} p^a(a_i|\zeta(s_i)) p(o_i|s_i, u_i) \\ &\quad \delta\left(\psi(s_j) - \min(\psi(s_i) + a_i - o_i u_i, b_L)\right), \end{aligned} \quad (15)$$

where function $\delta(x)$ returns 1 if $x = 0$ and 0 otherwise. Functions $\chi(s)$, $\psi(s)$, and $\zeta(s)$ give the channel state, buffer state, and incoming traffic state respectively of the composite state s. Let μ^n denotes a decision rule at time-slot n, then $\mu^n : \mathcal{S} \mapsto \mathcal{U}_{\mathcal{S}}$. The *policy* π specifies the decision rules to be used at all time-slots. A policy is called stationary if $\mu^n = \mu, \ \forall \ n \in \mathcal{T}$; for brevity we denote it by μ. The expected average cost per time-slot with stationary policy μ is

$$J_\mu = \lim_{m \to \infty} \frac{1}{m} \mathbb{E}\{\sum_{n=1}^{m} g(s^n, \mu(s^n))\}. \quad (16)$$

If the current system state is known perfectly, the optimal policy μ^* over all stationary policies that minimizes (16) can be found using dynamic programming (DP) algorithm for Markov decision Process (MDP). The corresponding Bellman equation can be given by [10],

$$\lambda + h(s_i) = \min_{u \in \mathcal{U}_{s_i}} [g(s_i, u) + \sum_{s_j \in \mathcal{S}} p_{s_i,s_j}(u) h(s_j)], \quad (17)$$

where $s_i = s_1, s_2, \cdots, s_N$, λ is the optimal cost, and $h(s_i)$ has the interpretation of a relative cost for state s_i.

Unlike MDP model, in a POMDP model the state of the system is not known exactly, so the previous approach cannot be applied directly. However, based on the observations and the actions taken, a *belief state* of the system can be formed. The belief state z is defined as a probability distribution over all possible states given the history of actions and observations. By maintaining the prior distribution over states, $p(s^{n-1})$ the belief can be computed recursively using Bayes rule. In the ARQ case, at the start of a particular time-slot, the observation obtained is for the action chosen in the previous time-slot. Therefore, we have to slightly modify the belief state update equation for POMDP problem of [11] by introducing the notion of initial and updated beliefs. For the problem at hand, the initial estimate of the belief at particular time-slot $n \in \mathcal{T}$ can be written as,

$$z^n(s_j) = \beta_1 \sum_{i=1}^{N} p_{s_i,s_j}(u^{n-1}) \tilde{z}^{n-1}(s_i), \ \forall s_j \in \mathcal{S}. \quad (18)$$

After getting the observation for certain action taken, the belief state at time-slot n for all $s_j \in \mathcal{S}$ can be updated as,

$$\tilde{z}^n(s_j) = \beta_2 p(o^n|u^n, s_j) \sum_{i=1}^{N} p_{s_i,s_j}(u^{n-1}) \tilde{z}^{n-1}(s_i). \quad (19)$$

In (18) and (19), β_1 and β_2 are normalizing constants that make the belief distribution sum to 1. Whereas a MDP policy dictates an action for every physical state, a POMDP policy dictates an action for every belief state. Again, it can be shown that solving a POMDP on a physical state is equivalent to solving a MDP on the corresponding belief state [11]. Since a POMDP can be considered as a belief state MDP [11], it may seem reasonable to apply DP algorithm for MDP for finding optimal policy over the belief state space directly. Unfortunately, the belief space is continuous which would mean applying DP algorithm to an uncountably infinite state space. Although there are algorithms that can solve POMDPs exactly, none of them are useful in solving our problem at hand. These algorithms apply to the finite horizon POMDPs with bounded costs and can be used for solving problem upto 5 states [12]. It has been shown that finding optimal policy even for a simplified finite horizon POMDP is PSPACE complete. On the other hand, there are a number of heuristic for finding policies approximately that perform well in a number of real world situations. In the sequel, we shall describe two of them and apply these heuristic for solving our problem.

A. *Maximum-likelihood Policy Heuristic:*

At a particular time-slot, maximum-likelihood (ML) heuristic finds the most probable state from the belief state of the system. Then the optimal policy for the chosen state is found from the DP algorithm for underlying MDP and is applied for that time-slot. At time-slot n, ML policy can be given as [13],

$$\mu_{ML}(z^n) = \mu^*_{MDP}(\arg\max_{s}\,(z^n(s))). \tag{20}$$

In this expression, $\mu^*_{MDP}(s_i)$ is the optimal policy for state $s_i = s_1, \cdots, s_N$ of the system and can be determined as,

$$\mu^*_{MDP}(s_i) \in \arg\min_{u \in U_{s_i}} \{g(s_i, u) + \sum_{s_j=s_1}^{s_N} p_{s_i,s_j}(u) h(s_j)\}, \tag{21}$$

where $h(s_j),\ s_j \in \mathcal{S}$ can be determined via relative value iteration algorithm for MDP using Bellman equation (17). The ML policy given by (20) is stationary since the optimal policy of the underlying MDP is stationary. It is intuitively clear that this policy can only perform well when the most probable state of system is much more probable than all the other states. Note that other good heuristical state-dependent policies can be used instead of μ^*_{MDP}.

B. *Voting Policy Heuristic:*

A major problem of ML heuristic is that it completely neglects all but a single state to determine the action. It chooses a particular action that is optimal for the most likely state despite the fact that the system is more likely to be in a state where other action is the best action. Instead of state, voting heuristic assigns a probability distribution over the actions. Each state votes for an action as determined by the optimal policy of the underlying MDP. At a particular time-slot, the vote is weighted by belief of the state and sum of all weighted votes for each action is determined. The action with largest sum is selected as the optimal action. As in ML case, during planning only the underlying MDP is solved using DP algorithm, however a full belief is maintained during execution. The voting heuristic policy at time-slot n can be obtained as [14],

$$\mu_{voting}(z^n) = \arg\max_{u} \sum_{s_i \in \mathcal{S}} z^n(s_i)\delta(\mu^*_{MDP}(s_i) - u) \tag{22}$$

IV. Simulation Results

In this section, we present some simulation results that illustrate the performances of the heuristic discussed in Section III. We also compare their performances with fully observable channel state (FOCS) case. In all simulations, we assume that the current buffer state is fully observable and the channel state is partially observable, but belief is known. The incoming traffic is Bernoulli distributed with packet arrival probability being q. We plot average throughput, average packet errors and average buffer delay curve as a function of packet arrival probability for different Doppler frequencies, different sets of actions, different transmission powers, different weighting factors, different number of channel states, and different buffer sizes. All curves are plotted for average channel power gain $\bar{\gamma} = 1$, noise power $\sigma^2 = 1$ mW, block duration $T_B = 255 \times 10^{-6}$ seconds, horizon $m = 10^5$, size of transmitted packet (block) $n' = 255$ bits, $8-$QAM transmission, and maximum number of packets arrival in a time-slot $A = 1$.

In Fig. 2, the effect of different number of actions on

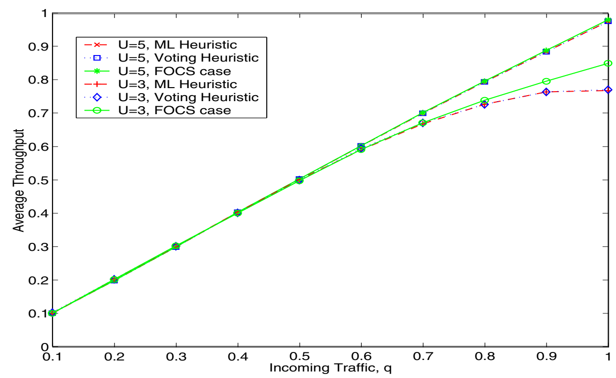

Fig. 2. Effect of number of actions on throughput

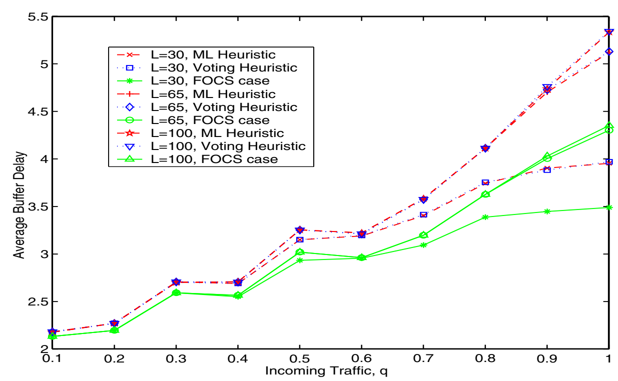

Fig. 3. Influence of buffer size on delay

the two heuristic and FOCS case is shown. The curves are plotted for number of channel state $C = 6$, weighting factor $\beta = 10$, buffer size $L = 30$ packets, transmission power $P_t = 4$ mW, Doppler frequency $f_m = 100$ Hz. We take the following actions: for $U = 3$, $u_1 =$ no transmission, $u_2 = (255, 123, 19)$ BCH code, $u_3 = (255, 247, 1)$ BCH code and for $U = 5$, $u_1 =$ no transmission, $u_2 = (255, 63, 30)$ BCH code, $u_3 = (255, 123, 19)$ BCH code, $u_4 = (255, 187, 9)$ BCH code, $u_5 = (255, 247, 1)$ BCH code. It is seen that the average throughput increases as U increases. Because as the value of U increases, the system possesses more flexibility to deal with the time-varying channel, which results in its improved performance. It is also seen from the curves and simulations

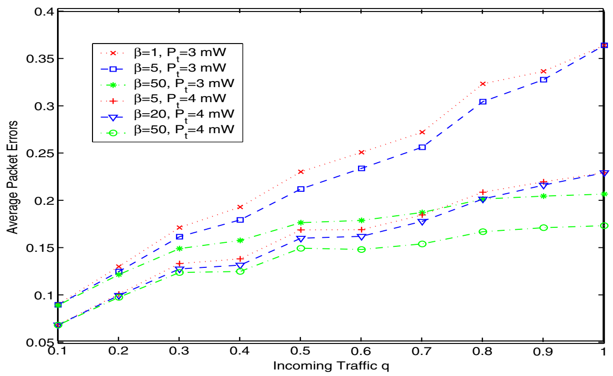

Fig. 4. Effect of β and P_t on average packet errors

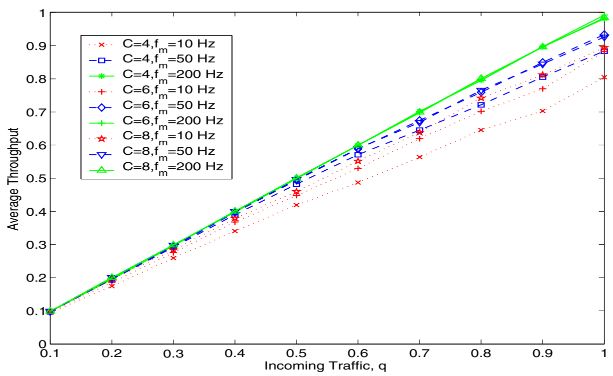

Fig. 5. Variation of throughput for different f_m and different C

that the performances of both ML and voting heuristic are close to FOCS case. Both heuristic perform better as U increases. The influence of buffer size on the average delay for all heuristic and FOCS case is shown in Fig. 3. Results are given for: $C = 6$, $f_m = 100$ Hz, $\beta = 10$, $P_t = 4$ mW, and $U = 5$ given previously. The curves show that the average delay increases as the buffer size increases. This fact can be explained by understanding the fact that as the buffer size increases the scheduler has more opportunity to store packets in the worse channel conditions. The scheduler sends these packets in good channel condition with less NAK probability. Therefore, the throughput increases, packet error decreases and delay increases with the increase of buffer size. The effect of weighting factor for two transmission powers is shown in Fig. 4. As β increases, the scheduler puts more importance on the packet error than delay, therefore, it transmits more cautiously so that the NAK probability is less. This in turn reduces number of retransmissions. Hence throughput increases and packet error decreases, but delay increases as the value of β increases. It is also to note that the packet error decreases when the transmission power increases. The numerical data for these curves are: $C = 6$, $f_m = 100$ Hz, $L = 30$ and $U = 5$ given for Fig. 2. In Fig. 5, we plot the throughput for different Doppler frequencies and different number of channel states. The data for the curves are: $L = 30$, $P_t = 4$ mW, $\beta = 10$ and $U = 5$ given for Fig. 2. It can be observed that the throughput increases as the Doppler frequency increases. This is due to the fact that when fading rate increases the chance of getting better channel condition soon after bad channel state is increased and thus packet storing and dropping probability decreases. This in turn increases throughput and decreases delay. The effect of C can be explained by understanding the fact that the value of r that depends on f_m and C should not be too large or too small [8]. Voting heuristic is considered for Fig. 4 and 5.

V. Conclusions

We have shown that finding optimal policies for coding rate adaptive HARQ system over time-varying channel is a POMDP problem. Since optimal solution is infeasible, this paper explores the effectiveness of two policy heuristic. The policies are function of belief of the states. Performances of these heuristic are evaluated in terms of average throughput, average delay and average packet errors and are found almost same as fully observable channel state case. Results show that throughput increases with number of actions, buffer length, weighting factor and Doppler frequency.

Acknowledgment

This work is supported jointly by the Natural Sciences and Engineering Research Council (NSERC) of Canada under a strategic project grant and UBC Graduate Fellowship award.

References

[1] A. Chockalingam, M. Zorzi, L. B. Milstein, and P. Venkataram, "Performance of a wireless access protocol on correlated Rayleigh-fading channels with capture," *IEEE Trans. Commun.*, vol. 46, pp. 644-655, May 1998.

[2] B. Vucetic, "An adaptive coding scheme for time-varying channels," *IEEE Trans. Commun.*, vol. 39, pp. 653-663, May 1991.

[3] M. Rice, and S. B. Wicker, "Adaptive error control over slowly varying channels," *IEEE Trans. Commun.*, vol. 42, pp. 917-925, Feb./Mar./Apr. 1994.

[4] S. S. Chakraborty, M. Liinabarja, and E. Yli-Juuti, "An adaptive ARQ scheme with packet combining for time varying channels," *IEEE Commun. Lett.*, vol. 3, pp. 52-54, Feb. 1999.

[5] H. Minn, M. Zeng, and V. K. Bhargava, "On ARQ scheme with adaptive error control," *IEEE Trans. Veh. Technol.*, vol. 50, pp. 1426 - 1436, Nov. 2001.

[6] Sunghyun Choi, and K. G. Shin, "A class of adaptive hybrid ARQ schemes for wireless links," *IEEE Trans. Veh. Technol.*, vol. 50, pp. 777-790, May 2001.

[7] C. C. Tan, and N. C. Beaulieu, "On first-order Markov modeling for the Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 48, pp. 2032-2040, Dec. 2000.

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

[9] M. K. Simon, and M. -S. Alouini, *Digital Communication over Fading Channels*. New York: John Wiley & Sons, 2000.

[10] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, Second Edition, vol.II. Belmont, Massachusetts: Athena Scientific, 2001.

[11] S. Russel, and P. Norvig, *Artificial Intelligence: A Modern Approach*. 2nd ed., New Jersey, NJ: Prentice Hall, 2003.

[12] M. Littman, A. Cassandra, and L. Kaelbling, "Learning policies for partially observable environments: scaling up," in *Proc. of the International Conference on Machine Learning*, San Francisco, CA, pp. 362-370, 1995.

[13] I. Nourbakhsh, R. Powers, and S. Birchfield, "DERVISH: An office-navigating robot," *AI Magazine*, vol. 16, no. 2, pp. 53-60, 1995.

[14] R. Simmons, and S. Koenig, "Probabilistic navigation in partially observable environments," in *Proc. of the International Joint Conference on Artificial Intelligence*, Montreal, Canada, pp. 1080-1087,Aug. 1995.