# A Stochastic Control Approach for Scheduling Multimedia Transmissions over a Polled Multiaccess Fading Channel

Munish Goyal, Anurag Kumar and Vinod Sharma

Dept. of Electrical Communication Engineering,

Indian Institute of Science, Bangalore, India.

{munish,anurag, vinod}@ece.iisc.ernet.in

## I. ABSTRACT

We develop scheduling strategies for carrying multimedia traffic over a polled multiple access wireless network with fading. We consider a slotted system with three classes of traffic (voice, streaming media and file transfers). A Markov model is used for the fading and also for modeling voice packet arrivals and streaming arrivals. The performance objectives are a loss probability for voice, mean network delay for streaming media, and time average throughput for file transfers. A central scheduler (e.g., the access point in a single cell IEEE 802.11 wireless local area network (WLAN)) is assumed to be able to keep track of all the available state information and make the scheduling decision in each slot (e.g., as would be the case for PCF mode operation of the IEEE 802.11 WLAN). The problem is modeled as a constrained Markov decision problem. By using constraint relaxations (a linear relaxation and Whittle type relaxations) an index based policy is obtained. For the file transfers the decision problem turns out to be one with partial state information. Numerical comparisons are provided with the performance obtained from some simple policies.

**Keywords:** scheduling over fading wireless channels, indexability and index policies, QoS in 802.11 wireless LANs.

## II. INTRODUCTION

We consider a home or office environment, where mobile stations (MSs) communicate with the external world through a wired access point (AP) (e.g., an AP in an IEEE 802.11 WLAN) as shown in Figure 1. Access to the Internet and the phone network is through a wired access link (e.g, DSL, T1-E1 or TV Cable;

see Figure 1). We assume that at least over the wireless interface, the voice is packetised. The TV receives streaming media over the wireless network; this could be broadcasts over the cable or it could serve programming off the media server (e.g., in the home setting, the media server could record programs while the family is away in the day time). Of course, Internet access from personal workstations or laptop computers would also be over the wireless local area network (WLAN). It is well known that the different types of traffic we wish to carry (i.e., voice, streaming media and file transfers) have different quality of service (QoS) requirements. The problem thus is to ensure that all the services being carried over the WLAN obtain their required quality of service (QoS), and the system capacity is efficiently utilised. The main difficulty in achieving this in the WLAN environment is the location dependent and time varying wireless channel conditions, or fading, and the limited availability of information regarding the system state.

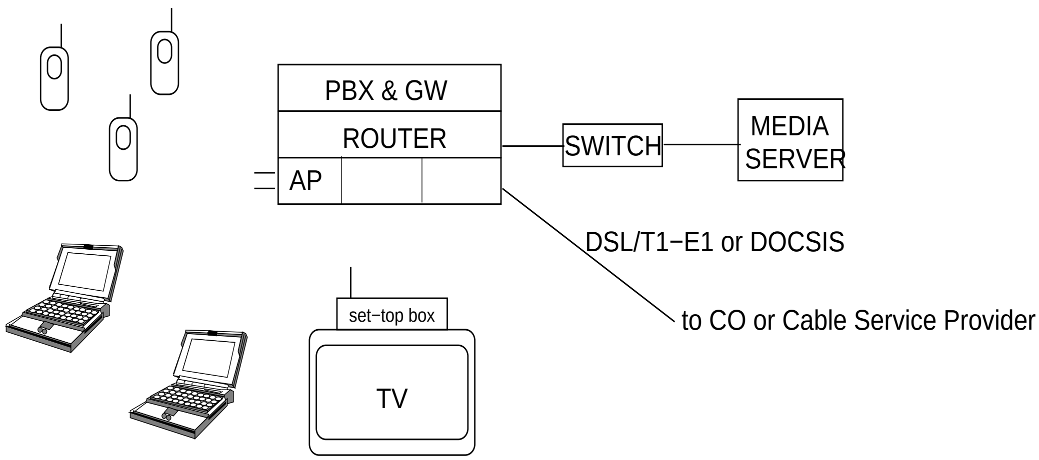



Fig. 1. A home or office wireless local area network being used for telephony, streaming media playback and Internet access.

All traffic will be assumed to be between the MSs and the AP. It is assumed that each MS has a separate virtual device for a voice, streaming or a file transfer session. The following are the parameters, models and performance objectives for each connection.

- *Packet Voice Telephony:* There are $N_V$ voice calls, each between an MS and the AP. We assume on-off model for voice and a voice call, when active, produces periodic packets. Letting $D_V$ be the (random) voice packet delay for a connection, the performance requirement is $Pr(D_V > T_V) < \epsilon_V$, where $T_V$ is a delay bound (e.g., 30ms), and $\epsilon_V$ is a small probability (e.g., 0.01). Packets that exceed their delay target are assumed to be lost. Though delays of the order of $150$ ms are tolerable, we assume that if the packet is delayed by more than $T_V$ at MS or AP, it is going to exceed 150 ms till it reaches the destination due to other network delays. We associate a cost, representing the number of packets dropped due to violation of delay constraint, with each call and design policies to minimize a long run average cost.
- *Streaming Media:* There are $N_S$ streaming multimedia connections (e.g., video or audio). We will

assume that a streaming media source generates packets according to a Markov process. Streaming traffic can be buffered at the receiver for smooth playout, and the amount of buffering can be substantial since the interactivity requirements are not particularly strict. When playing out a movie from a server (see Figure 1) the user may wish to stop, fast forward or rewind. If excessive packets from a movie are buffered in the AP and if a user command necessitates new packets be brought in from the server then the queued packets will add to the command response time resulting in an annoying behavior. Thus we associate with each streaming connection a holding cost indicative of the number of packets buffered at the source. First, we look at the discounted packet holding cost with a discount factor $\alpha \in (0,1)$ and then in the limit as $\alpha \to 1$, this discount holding cost is equivalent to the mean queueing delay by Little's law. The mean queueing delay requirement for streaming traffic is $d_S$.

- *File Transfers:* There are $N_T$ file transfers between the wired network and the MSs via the AP. These will be taken to be large volume transfers. We are therefore interested in the throughput of such transfers, and this will be denoted by $\sigma_T$. We associate a throughput reward with each session and wish to maximize a long run average reward.

In this paper we assume that a polling station (PS) (collocated with the AP) provides centralized, contention-free channel access, based on a poll-and-response mechanism. A virtual connection is established before commencing a transfer requiring some parameterized quality of service (QoS). A set of traffic characteristics are negotiated between the AP and the corresponding station. Accordingly, the AP implements an admission control algorithm to determine whether to admit a specific connection or not. Once a connection is set up, the PS endeavors to provide the contracted QoS by allocating the required resources. In order to meet the contracted QoS requirements, the PS needs to schedule the data and poll frame transmissions. Since the wireless medium involves time-varying and location-dependent channel conditions, developing a good scheduling algorithm is a challenging problem. A well designed scheduling algorithm can result in better system performance, i.e., more traffic can be handled for given QoS requirements (See Figure 2). In a typical frame exchange sequence, the PS polls a station asking for a pending frame. If the PS itself has pending data for this station, it uses a combined data and poll frame by piggybacking the poll frame into the data frame. Upon being polled, the polled station acknowledges the successful reception of the frame sent by the PS along with data asked for by the PS. The PS then polls the next station as prescribed by the scheduling algorithm based on the current system state.

With the above situation in mind we consider a model with periodic frames of equal length. The polling decisions would be taken at the start of each frame. For each connection, there would be a queue at the corresponding MS and a queue at the AP side. For each voice connection, one packet is generated per frame during active period. The packet arrival model for streaming traffic is a Markov process embedded at the

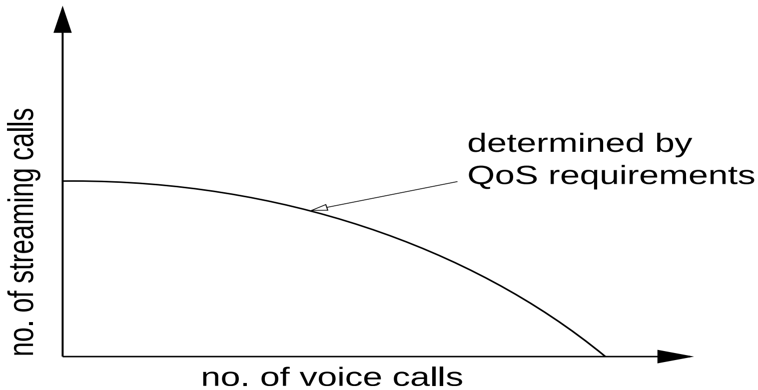



Fig. 2. The schedulable region for voice and streaming media calls. For each point in the region the resources can be allocated or scheduled among that many traffic flows so as to meet QoS objectives for each connection. For a given system, given traffic characteristics, and given performance objectives, the network should operate in a way that makes the schedulable region as large as possible.

frame boundaries. The file transfers are assumed to have backlogged data. In the queue on the side of the file source, i.e., if the MS is downloading a file then the queue at the AP is backlogged, whereas if the MS is uploading a file then the queue at the MS is backlogged. The channel gain between any transmitter-receiver pair is constant over each frame but varies in a Markovian manner from frame to frame. We assume that the channel gain seen during transmission from the AP to an MS is same as the one seen for transmission from that MS to the AP in the same frame; this channel *reciprocity* is valid since the communication is time division duplex and hence the transmissions both ways take place at the same frequency. In this framework, our aim is to develop dynamic scheduling policies that optimize certain long run performance objectives. A long run performance objective do make sense as the call durations for the traffic classes under consideration are fairly long. We model the system mathematically and analyse it using the dynamic programming approach.

The frame would be divided into three subframes; one for each traffic class (see Figure 3). Since the channel is time varying, the actual time taken for transmission and hence the length of a subframe varies. We introduce bounds on the minimum and the maximum time available for each subframe. These bounds could then be tuned to satisfy the above said quality of service constraints. Note that this does not limit the generality of the problem, since for example, we may say that all subframe lengths are upper bounded by the frame length itself. There would be a priority order, with voice calls given the highest priority whereas the file transfer traffic given the least. This is justified since the voice packets cannot be stored beyond $T_V$, streaming packets cannot be stored for long and the file transfer traffic normally uses the available bandwidth. Choosing a lower bound for the lowest priority traffic subframe length would provide a lower bound on its performance. The time left over by a subframe of a higher priority class can be used by a lower priority class.

The model discussed above has been widely considered in DOCSIS networks [2] and other TDMA based networks such as satellite networks. In [5], Capone and Stavrakakis have considered a problem of designing admission control and scheduling algorithms for time-division multiple access wireless systems supporting variable bit rate applications. The quality of service is expressed in terms of tolerable delay. Fading was not considered in the model. Similar problem has been looked at in [9] for DOCSIS networks. Recently, there has been a lot of interest in delay optimal scheduling of transmissions over fading wireless networks [4], [7], [15]. The optimal policies more often than not turn out to be too complicated. The major contribution of this work is the development of index based polling strategies. This paper is organized as follows. In Section III, we model the system under consideration. We formulate the problem mathematically in Section IV. We obtain polling strategy for the voice calls in Section V. We consider the performance optimization problem for streaming calls in Section VI followed by a formulation of a relaxed version of the problem in Section VI-B. This is followed by a detailed analysis of the relaxed problem using the dynamic programing technique. An index based heuristic polling policy for streaming calls is obtained in Section VI-E. We obtain an index policy for file transfers in Section VII.

## III. System Model

Let there be a set $\mathcal{N}$ of virtual devices in the system. Time is divided into fixed length frames of duration $\tau$ seconds each. The frame is divided into three subframes, one per class. The subframe length for the voice class is upper bounded by $\tau_V$ and that for the file transfers is lower bounded by $\tau_T$. The subframe length for the streaming traffic is thus upper bounded by $\tau - \tau_T$. See Figure 3 for details. Voice traffic is given the highest priority whereas the file transfers are given the least priority subject to the above subframe length constraints. A voice connection $i \in \mathcal{N}_V$, when active, generates a packet of size $b_i$ per frame. A packet generated during frame $n$ can only be sent in frame $n+2$ and if not sent in that frame it is considered lost; this bounds the voice packet delay to three times the frame time. A streaming connection $i \in \mathcal{N}_S$ (for example a variable rate coded video source) places a random number of packets, each of length $b_i$, into its transmitter buffer (of infinite capacity) at the start of each frame. We assume that the packet arrival process $A_i[n]$ is a finite state Markov chain with a single ergodic class and the transition probability matrix is $\mathbf{P}_i^{(a)}$ for $i \in \mathcal{N}_S$. The source side queue of a file transfer connection $i \in \mathcal{N}_T$ has infinite backlog of packets to be sent (this could be the case if the file transfers are window controlled with large window as in TCP).

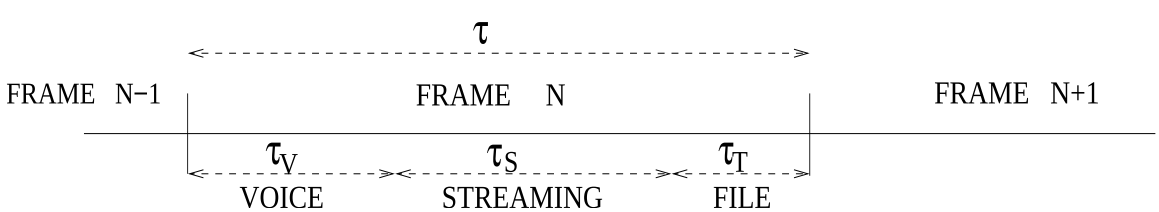



Fig. 3. A typical frame showing voice, streaming and file transfer subframes.

A link is defined as a source and sink pair. The channel "power" gain process for a link is assumed to remain constant over the duration of a frame and is modeled as a finite state Markov chain with a single ergodic class, embedded at the frame boundaries, with transition probability matrix $\mathbf{P}_i^{(h)}$ for link $i$. The channel gain process is assumed to be independent from one link to another. Note that the channel is reciprocal. A peak power constraint is generally imposed for all devices in a wireless environment (as in the IEEE 802.11 standard). Based on the link gains, we can compute a maximum reliable transmission rate for each device when transmitting at this peak power level. This is done using a well known mapping between signal to noise ratio and the transmission rate for reliable transmission. Let $R_i[n]$ be the transmission rate, in terms of packets per second, from node $i$ during frame $n$. It follows that the process $R_i[n]$ for transmitter $i$ is also a finite Markov chain with transition matrix $\mathbf{P}_i^{(r)}$. For simplicity, we assume that $R_i[n]$ is strictly positive for all $i$. See Figure 4

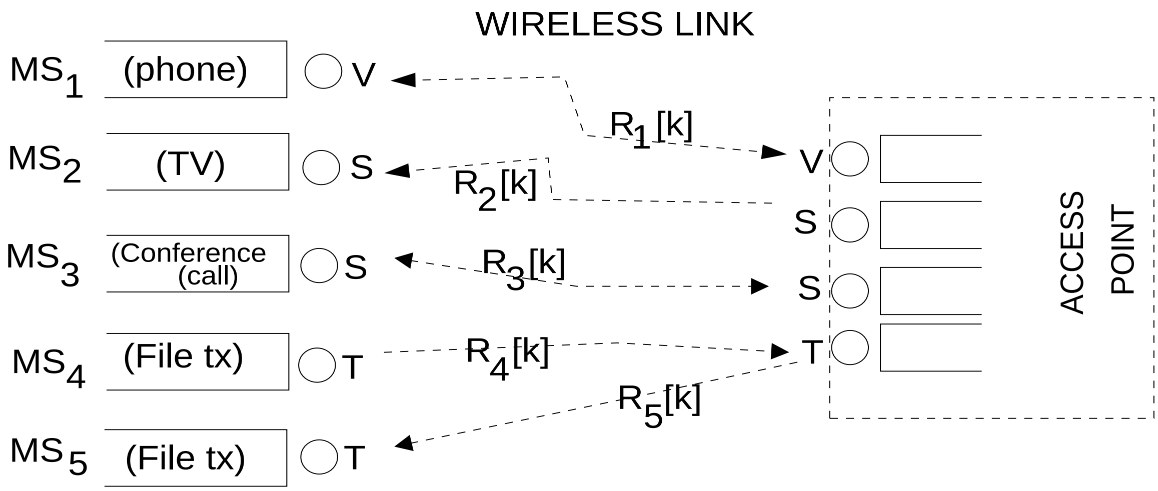



Fig. 4. A typical configuration of a wireless local area network. $MS_1$ carries voice (V) connection on a link 1 and the rate process is $R_1[k]$. $MS_2$ plays a movie (streaming connection S) off the media server on link 2 and rate process is $R_2[k]$. $MS_3$ is engaged in a conference call over the network (naturally a bidirectional transfer) using link $R_3[k]$. $MS_4$ and $MS_5$ are uploading and downloading files (shown T) over the Internet. Thus $\{MS_1\} \in \mathcal{N}_\mathcal{V}$, $\{MS_2, MS_3\} \in \mathcal{N}_\mathcal{S}$ and $\{MS_4, MS_5\} \in \mathcal{N}_\mathcal{T}$.

At time instant $n\tau, n = \{0, 1, 2 \cdots\}$, the AP is provided with the information about the available transmission rates $R[n]$ for all links that carry streaming and voice traffic. The information regarding the number of packets $A[n]$ that arrive during the previous frame is also provided to the AP. Thus the AP would know the buffer lengths at each streaming source and which of the voice sources have packet to send. We propose to introduce a field in the packet header to convey the information. In a recent draft of IEEE 802.11e, a field carrying the queue length information has already been added. A streaming or a voice source which is not scheduled to transmit during a frame will also be polled to get the current information regarding the transmission rates and the arrivals. Such a device will not send any data upon being polled except that the header bits are set appropriately in the response packet to convey the desired information (e.g, use CF-Poll+CF-Ack (no data) type frame (See [1])). Arrival information during frame $n$ will be communicated to AP during frame $n+1$ and the decision process would include these packets while making polling decisions for the frame $n+2$. Since the number of streaming and voice sessions are small in number as they are admission controlled, this way of polling each device is reasonable. But for file transfer sessions, of which there are many, the exchange of these null packets could be waste of time. Thus we assume that only partial (delayed) information is available regarding the available transmission rate for a link carrying a file transfers. For such a session the AP knows the transmission rate at which the last transmission from that source occurred and the time since last transmission. Thus, gives a probability measure over the channel transmission rates.

Based on the available information, the AP decides upon a subset of devices that can send and how much they can send in the current frame, i.e., during the time period $[n\tau, (n+1)\tau)$. The objective of the AP, which acts as a controller, is to obtain an optimal resource (frame time) allocation or polling strategy that guarantees a desired quality of service for each device subject to the constraints imposed by the wireless network. This policy would yield a schedulable region comprising of sets $\mathcal{N}_V$ and $\mathcal{N}_S$ which can be handled by the system so that each session obtains its desired QoS. Given that the number of admitted voice and streaming calls belong to this region, we can find the maximum throughput available for the file transfer traffic.

## IV. Problem Formulation

We associate with device $i \in \mathcal{N}$, a weight $\omega_i$ defining its priority over other devices. The voice call is a two way communication. For example, there will be two packets generated per frame for each such call, one at the MS and the other at the AP, if both sides are active. By reciprocity of the channel, we can view it as one device with two packets to be transmitted per frame and the channel gain is the gain of the link over which the call is handled. The number of voice packets generated per frame for a voice call $i \in \mathcal{N}_V$ is $Q_i[n] \in \{0, 1, 2\}$; let $S_i[n] \leq Q_i[n]$ be the number of packets transmitted in the $n^{th}$ frame at a rate $R_i[n]$, i.e., during time $[n\tau, (n+1)\tau)$, where $n = \{0, 1, 2, \cdots\}$. If $S_i[n] = 1$ and $Q_i[n] = 2$, one can choose to transmit any one of the two packets as the frame cost would be the same. The objective of minimizing the packet loss probability is captured by maximizing the expected number of packets transmitted. Given $Q_i, R_i$ for $i \in \mathcal{N}_V$, the controller objective is to maximize a weighted sum of the expected number of packets transmitted subject to the subframe length constraint,

$$\max_{\{S_i \leq Q_i, i \in \mathcal{N}_V\}} \left\{ \sum_{i \in \mathcal{N}_V} \omega_i S_i : \sum_{i \in \mathcal{N}_V} \frac{S_i}{R_i} \leq \tau_V \right\}. \tag{1}$$

Based on the optimal actions above, let $T_V[n]$ be the time occupied by voice packets in frame $n$. Next we consider a streaming device $i \in \mathcal{N}_S$. Again there could be two queues per streaming call, one at the MS and other at the AP. By reciprocity we can look at it as a single queue associated with the MS and the channel gain seen for the transmission would be the link gain between the MS and the AP. If the solution turn out to be to serve say $s$ packets in frame $n$ for MS $i$, then how many packets would be served from each of the two queues can be defined arbitrarily as the cost would be the same (longest queue first policy may be reasonable). Thus from analysis point of view, the two situations, first being that of two queues one at the AP and other at the MS and second being a single queue at the MS with aggregate arrival process, are equivalent. Let $A_i[n]$ be the number of packets that arrive during $[(n-1)\tau, n\tau)$ (see Figure. 5). Note that it is the sum of those arrived at the MS side and those at the AP side. Arriving packets are placed into the transmitter buffer at the end of each frame. Let $Q_i[n]$ be the queue length at time instant $n\tau$ for device $i$. Let $S_i[n]$ be the number of packets transmitted in the $n^{th}$ frame, i.e., during $[n\tau, (n+1)\tau)$. Obviously, $S_i[n] \in [0, Q_i[n]]$, since one can transmit only up to whatever is available in the buffer. The transmitter queue evolves according to the equation $Q_i[n+1] = Q_i[n] - S_i[n] + A_i[n]$ (see Figure 5).

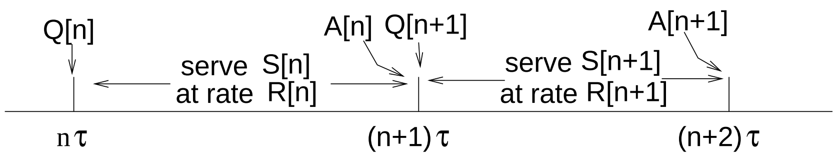



Fig. 5. Model for service to a streaming transfer

Focusing only on the streaming transfers, the quadruplet $\mathbf{X} = (\mathbf{Q}, \mathbf{R}, \mathbf{A}, T_V)$ defines the state of the system, where $Q$ is the queue length and the $R$ is the transmission rate available. The quality of service measure is $\sum_{k=0}^{\infty} \alpha^k Q_i[k]$, where $\alpha \in (0,1)$ is a discount factor. If $\alpha$ is small, the recent queue lengths have more value than those in a distant future whereas if $\alpha$ is large, queue lengths in a distant future are also important. The maximum subframe length available for streaming traffic is $\tau - \tau_T - T_V[k]$. The controller objective is to obtain a sequence $\{S_i[k]\}, i \in \mathcal{N}_S$ that minimizes a weighted sum of the performance measure subject to the subframe length constraint,

$$\min \left\{ \sum_{i \in \mathcal{N}_S} \omega_i E\Big[ \sum_{k=0}^{\infty} \alpha^k Q_i[k] \Big] : \sum_{i \in \mathcal{N}_S} \frac{S_i[k]}{R_i[k]} \leq \tau - \tau_T - T_V[k];\ S_i[k] \in \{0, 1, \cdots, Q_i[k]\};\ \forall k \geq 0 \right\}, \quad (2)$$

where the measure over which the expectation operator $E$ is taken is conditioned on the state at time $k = 0$, and the actions $\mathbf{S}[k] = \{S_i[k], i \in \mathcal{N}_S\}$ are based on the history of the process. This is a Markov Decision process with state dependent action space and a hard constraint in each step. Recall that the sequence of actions $S_i[k]$ are integer valued. As $\alpha \to 1$, the control actions would minimize the mean packet transmission delay.

Based on the optimal actions for streaming and voice traffic, let $T_S[k]$ be the time occupied by a streaming traffic during the $k^{th}$ frame. The available subframe length for file transfers is $T[k] := \tau - T_V[k] - T_S[k]$. Note that the lower bound on the subframe length for such traffic is satisfied. The channel state of the link over which the file transfer traffic is carried is known at the AP only when the transmission actually takes place. Since there is a large number of such sessions we would not be able to poll all devices with dummy packets as we did for streaming traffic. The controller instead can keep track of the rate at which the last transmission for a particular session took place and the delay in terms of the number of frames since the last transmission. Thus for each such connection, in any slot this state yields a probability distribution on the available transmission rate. Let, at the beginning of the frame $k$, $r_i$ be the rate at which the last transmission took place for connection $i$ and $d_i$ be the number of slots since the start of the last transmission. Thus at time instant $k$, the probability distribution is $\pi(r) := P_{r_i}^{d_i}(r)$. Thus we define the system state as a vector $(\mathbf{r}, T, \mathbf{d})$. Let $\mathcal{S}_i$ be the space of all possible pairs $(r_i, d_i)$.

Let $S_i[k]$ be the action, representing the number of packets transmitted by a file transfer session during $k^{th}$ frame. The state evolution equation is given by,

- $r_i[k+1] = r_i[k]$ if $S_i[k] = 0$
- $r_i[k+1] = j$ if $S_i[k] > 0$ and the packet is transmitted at rate $j$
- $d_i[k+1] = d_i[k] I_{\{S_i[k]=0\}} + 1$

Given the state vector $\mathbf{x} = (\mathbf{r}, T, \mathbf{d})$, an action $S_i$ yields a reward $\omega_i S_i$. The constraint on the subframe length $T[k]$ should be satisfied. The objective is to obtain the policy $S_i[k]$ that would maximize the average reward while the subframe boundary constraint is not violated.

## V. Analysis: Voice Calls

First, we consider the problem stated in Equation (1). This problem is identical to a knapsack problem where there are certain quantities of material of different densities, and different sizes having different associated values per unit quantity. The number of items need to be chosen to fit into a container while maximizing the aggregate value. During the $n^{th}$ frame, the knapsack volume is the subframe time $\tau_V$, the transmission time per packet for the $i^{th}$ call is $\frac{1}{R_i[n]}$ and the value per packet associated with the $i^{th}$ call is $\omega_i$. The following is a well known heuristic for the above said problem obtained from a linear relaxation of the integer knapsack problem [6].

Order the devices in decreasing order of $\omega_i R_i[n]$; this can be interpreted as the reward per unit transmission time for device $i$. Determine $m_V[n]$ so that the $(m_V[n]+1)^{th}$ queue in this order can send at most one packet without violating $\tau_V$, the subframe length constraint. Now, for a queue $i$ among the top $m_V[n]$ queues in this order $S_i[n] = Q_i[n]$, and $S_i[n] = 0$ for the rest. The $(m_V[n]+1)^{th}$ queue can send at most one packet if possible. We could have sent a fraction of the packet at $(m_V[n]+1)^{th}$ queue but this would violate our modeling assumption that a packet cannot be fragmented. This policy yields a schedulable region for the voice calls determined by the QoS requirements. Define $T_V[k]$ the subframe time used by the voice traffic in the $k^{th}$ frame and is given by

$$T_V[k] = \sum_{i=1}^{m_V[k]} \frac{Q_i[k]}{R_i[k]} + \frac{I_{\{S_{m_V[k]+1}[k]=1\}}}{R_{m_V[k]+1}[k]}.$$

# VI. Analysis: Streaming Transfers

In view of the above result, the problem stated in Equation 2 can be restated as follows. For notational ease, we denote the random process representing the frame time available for streaming transfers $\tau - \tau_T - T_V[k]$ by $T[k]$. A realization of $T[k]$ will be denoted by $t$. Note that the process $T[k]$ is a Markov chain with finite state space since $\tau_V[k]$ can assume only finitely many values. The state of the system is now a quadruplet $\mathbf{X} = (\mathbf{Q}, \mathbf{R}, \mathbf{A}, T)$. The controller objective is to obtain a sequence $\{S_i[n]\}, i \in \mathcal{N}_S$ that solves

$$\min \sum_{i \in \mathcal{N}_S} \omega_i E\Big[\sum_{k=0}^{\infty} \alpha^k Q_i[k]\Big], \text{ subject to, } \sum_{i \in \mathcal{N}_S} \frac{S_i[k]}{R_i[k]} \leq T[k]; \; S_i[k] \in \{0, 1, \cdots, Q_i[k]\}, \; i \in \mathcal{N}_S \tag{3}$$

Using a heuristic based on the MDP formulation, we obtain an index based polling policy. A policy that orders the transmissions in decreasing order of $\{\omega_i r_i q_i\}$ is known to be stabilizing [3] for such a system. Note that this policy is also an index policy. It should also be noted that while the property of being stabilising is essential, not every stabilizing policy will perform well in terms of the objectives in Equation 3. We will compare the performance of the stabilizing policy with that of the index policy that we would obtain based on MDP formulation.

## A. *Index Policies and Whittle's Relaxation*

Let us look at the discounted cost value iteration algorithm for solving the problem (Equation 3) to motivate the approach that we will follow in the rest of the paper. For a given state $\mathbf{x} = (\mathbf{q}, \mathbf{r}, \mathbf{a}, t)$, define the constraint set $S(\mathbf{x}) = \{\mathbf{s} : \mathbf{s} \in [0, \mathbf{q}]; \;\; \sum_{i \in \mathcal{N}_S} \frac{s_i}{r_i} \leq t\}$. Let $V(\mathbf{x})$ be the optimal expected discounted cost when starting in state $\mathbf{x}$. Consider the following value iteration algorithm,

$$V_{n+1}(\mathbf{x}) = \min_{\mathbf{s} \in S(\mathbf{x})} \{ \sum_{i \in \mathcal{N}_S} \omega_i q_i + \alpha E_{\mathbf{a},\mathbf{r},t}[V_n(\mathbf{q} - \mathbf{s} + \mathbf{A}, \mathbf{R}, \mathbf{A}, \mathbf{T})]\}.$$

where $E_{a,r,t}[\cdot]$ denotes the conditional expectation with respect to the arrival, the rate and the available time processes and $V_n(\mathbf{x})$ is a sequence of value function which will be later shown to converge to $V(\mathbf{x})$. Let $f_n$ be the optimal policy for the $n^{th}$ stage problem. Initialize $V_0(\mathbf{x}) = 0$. This implies $V_1(\mathbf{x}) = \sum_{i \in \mathcal{N}_S} \omega_i q_i$. Thus $f_2(\mathbf{x})$ is $\arg\min_{\mathbf{s} \in S(\mathbf{x})} \{ \sum_{i \in \mathcal{N}_S} \omega_i (q_i(1+\alpha) - \alpha s_i + \alpha E_{a_i}[A])\}$. This is a knapsack problem. Using Lagrangian approach, we associate a multiplier $\beta$ and thus $f_2(x, \beta)$ equals $\arg\min_{\mathbf{s} \in [0,\mathbf{q}]} \{\sum_{i \in \mathcal{N}_S} \beta \frac{s_i}{r_i} - \omega_i \alpha s_i\}$. The knapsack heuristic solution is $f_2(x, \beta)|_i = q_i \theta_i(r_i, \beta)$, where $\theta_i(r_i, \beta) = I_{\{\omega_i \alpha r_i \geq \beta\}}$. The parameter $\beta$ solves for the frame boundary constraint. In other words the solution is to order the users in decreasing order of $\omega_i r_i$ and the user with highest index transmits until the frame boundary constraint is exceeded or there is no data for transmission. This is an index policy. The index $\omega_i r_i$ is essentially that value of $\beta$ at which the system makes a transition from an active action ("send something") to passive action ("send nothing"); i.e., if $\beta > \omega_i r_i \alpha$ then $\theta_i(r_i, \beta) = 0$ and $\theta_i(r_i, \beta) = 1$ otherwise.

The function $V_2(\mathbf{x})$ is too complex to carry out any further iteration. Moreover, we are interested in index based policies similar to the one obtained for the voice calls because of their ease in implementation. There has been much work on obtaining index based policies for bandit problems. For multiarmed bandit problems, it is well known that the policies based on Gittin's indices are optimal [12]. Gittin showed that to each project one could associate an index $\nu_i(x_i)$, a function of the project $i$ and its state $x_i$ alone, and that the optimal policy is to operate the one with the largest index.

Consider the "restless bandits" problem of designing an optimal sequential resource allocation policy for a collection of stochastic projects (say $M$), each of which is modeled as a Markov decision chain having two actions at each state with associated rewards; an active action, which corresponds to engaging the project, and a passive action, which corresponds to letting it go. The passive projects can change state, in general through a given transition rule and hence the word "restless". A fixed number of resources needs to be allocated; i.e., at each time instant a fixed number of projects (say $k$) are active. The performance objective is to maximize the time-averaged reward rate. Whittle [14] presented a simple heuristic based on a tractable optimal solution to a relaxed version, where instead of requiring that $k$ projects be active at any time, $k$ *projects are needed to be active on average*. This yielded an upper bound on the optimal reward. Further the heuristic policy is a priority index rule associated with each project, that engages the top $k$ projects at any given point of time. The recent work of Nino-Mora [10] is nearly a complete reference for restless bandit problems.

Motivated by the Whittle's work on restless bandits, we introduce a relaxed problem. The state of the system is denoted by $\mathbf{x} = (\mathbf{q}, \mathbf{r}, \mathbf{a}, t) \in \mathcal{X}$. The set of feasible actions in state $\mathbf{x}$ is $S(\mathbf{x}) = [0, \mathbf{q}]$. Let $\Pi$ be the space of all feasible policies. A deterministic, stationary Markov policy $f \in \Pi$ is a measurable mapping from $\mathcal{X}$ to $[0, \mathbf{q}]$. For every $\beta > 0$, the Lagrange multiplier, define a cost function $c_\beta(\mathbf{x}, \mathbf{s}) = \sum_{i \in \mathcal{N}_S}(\omega_i q_i + \beta \frac{s_i}{r_i})$. The term $\beta \frac{s_i}{r_i}$ can be seen as a relaxed frame boundary constraint. The Lagrange multiplier $\beta$ has an economic interpretation. The value $\beta \frac{s_i}{r_i}$ is a penalty for transmitting more data and thus reducing the frame time possibly available for other connection. There is a trade off. If more data is sent for a connection that connections queue reduces but the connection is penalised for doing so. Obviously, the penalty increases with $s_i$. The relaxed problem is to obtain a policy $\pi \in \Pi$ that minimizes the expected discounted cost $E^\pi_{\mathbf{x}}[\sum_{k=0}^\infty \alpha^k c_\beta(\mathbf{X}[k], \mathbf{S}[k])]$. Note that the relaxed problem is separable. Thus we solve it for each connection $i$. The amount of data $s_i$ that can be transmitted in a frame of length $t$ should satisfy $\frac{s_i}{r_i} \leq t$, the residual frame boundary constraint. We drop the subscripts $i$. Without loss of generality assume that $\omega = 1$. Exploiting the separability, the relaxed problem (RP) for each user is

$$V(x) = \min_\pi E^\pi_x \left[ \sum_{k=0}^\infty \alpha^k \left( Q[k] + \beta \frac{S[k]}{R[k]} \right) \right], \text{ subject to, } S[k] \in \{0, 1, \cdots, Q[k]\}, \ \frac{S[k]}{R[k]} \leq T[k], \ \forall k.$$

Note that we have relaxed the sum constraint but not the individual constraint. The same problem holds for each user. We now analyse this per user problem in order to obtain certain indices.

### B. *Analysis of the Relaxed Problem*

The state $x$ is the quadruple $(q, r, a, t)$. Our model satisfies the nominal conditions (see [11], Proposition 2.1) required for the existence of the discount optimal stationary policy, and the value function $V(x)$ is obtained as a solution to the following dynamic programming optimality equation. Define $u = q - s$ and $U(x) = \{u \text{ integer} : (q - tr)^+ \leq u \leq q\}$. The variable $u$ is the residual number in the queue after the policy has acted in an interval. Then

$$V(q, r, a, t) = \min_{u \in U(x)} \{q(1 + \frac{\beta}{r}) - \beta\frac{u}{r} + \alpha E_{a,r,t} V(u + A, R, A, T)\}. \quad (4)$$

Define $H(u, r, a, t) = E_{a,r,t} V(u + A, R, A, T)$.

*Theorem VI.1:* $V(u, r, a, t)$ and hence $H(u, r, a, t)$ is convex nondecreasing in $u$.

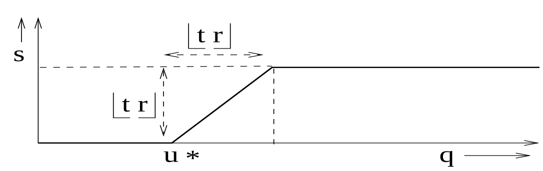
*Proof:* See the Appendix. ■

The unconstrained minimizer $u^*(r, a, t)$ in (4) is the value of $u$ that solves the following inequalities,

$$H(u, r, a, t) - H(u - 1, r, a, t) \leq \frac{\beta}{r\alpha} \leq H(u + 1, r, a, t) - H(u, r, a, t).$$

Note that the unconstrained minimizer is not a function of $q$. The solution for the constrained problem ($u \in U(x)$) is,

- $s(x) = 0$ for $q < u^*(r, a, t)$,
- $s(x) = \lfloor tr \rfloor$ for $q > u^*(r, a, t) + \lfloor tr \rfloor$,
- $s(x) = q - u^*(r, a, t)$ otherwise.




Observe that $u^*(r, a, t) = q$ is the break point that will be used to define the indices as in [14] as it is the boundary between not sending anything from the queue and sending something.

### C. *An Algorithm for Computing $u^*(\cdot)$*

Consider the discounted cost value iteration algorithm corresponding to the relaxed problem (4).

$$V_n(q, r, a, t) = \min_{u \in S(q,r,a,t)} \left\{q(1 + \frac{\beta}{r}) - \beta\frac{u}{r} + \alpha E_{a,r,t} V_{n-1}(u + A, R, A, T)\right\} \quad (5)$$

It follow from the proof of Theorem VI.1 that the functions $H_n(u, r, a, t)$ are convex in $u$ for each $n$. Let $u_n^*(r, a, t)$ be the value of $u$ that solves the following inequalities,

$$H_n(u, r, a, t) - H_n(u - 1, r, a, t) \leq \frac{\beta}{\alpha r} \leq H_n(u + 1, r, a, t) - H_n(u, r, a, t).$$

Based on the above said constrained solution, we have,

- If $q \leq u_n^*(r,a,t)$, $V_{n+1}(q,r,a,t) - V_{n+1}(q-1,r,a,t) = 1 + \alpha(H_n(q,r,a,t) - H_n(q-1,r,a,t))$
- If $u_n^*(r,a,t) < q \leq \lfloor tr \rfloor + u_n^*(r,a,t)$, $V_{n+1}(q,r,a,t) - V_{n+1}(q-1,r,a,t) = 1 + \frac{\beta}{r}$
- If $q > u_n^*(r,a,t) + \lfloor tr \rfloor$,

$$V_{n+1}(q,r,a,t) - V_{n+1}(q-1,r,a,t) = 1 + \alpha(H_n(q - \lfloor tr \rfloor, r,a,t) - H_n(q - \lfloor tr \rfloor - 1, r,a,t))$$

Define $W_n(q,r,a,t) = V_n(q,r,a,t) - V_n(q-1,r,a,t)$. Thus $H_n(q,r,a,t) - H_n(q-1,r,a,t) = E_{a,r,t}W_n(q + A,R,A,T)$. Then the iterative algorithm to compute $u^*(r,a,t)$ is as follows. Initialize $W_0(q,r,a,t) = 0$. Let $u_n^*(r,a,t)$ be the value of $u$ that solves the following inequalities,

$$E_{a,r,t}W_n(u + A,R,A,T) \leq \frac{\beta}{\alpha r} \leq E_{a,r,t}W_n(u+1+A,R,A,T). \tag{6}$$

The following procedure then obtain $W_{n+1}(\cdot)$ from $W_n(\cdot)$ and $u_n(\cdot)$.

- If $q \leq u_n^*(r,a,t)$, $W_{n+1}(q,r,a,t) = 1 + \alpha E_{a,r,t}W_n(q + A,R,A,T)$.
- If $u_n^*(r,a,t) < q \leq \lfloor tr \rfloor + u_n^*(r,a,t)$, $W_{n+1}(q,r,a,t) = 1 + \frac{\beta}{r}$.
- If $q > u_n^*(r,a,t) + \lfloor tr \rfloor$, $W_{n+1}(q,r,a,t) = 1 + \alpha E_{a,r,t}W_n(q - \lfloor tr \rfloor + A,R,A,T)$.

$u_{n+1}^*(\cdot)$ is thus calculated from Equation 6. The convergence of the value iteration algorithm (5) ensures that this algorithm converges and $u_n^*(r,a,t)$ converges to the optimal solution $u^*(r,a,t)$.

### *D. Indexability*

*Definition VI.1:* (Indexability) [14]: The system is said to be indexable if the set of states where a passive action is taken increases monotonically from an empty set to the full set as the parameter $\beta$ increases from $0$ to $\infty$. ∎

For our problem the requirement is natural. As the penalty $\beta$ for using the frame time increases, we choose to transmit less and less. We show that the relaxed problem is indexable in the sense of the above definition and obtain indices associated with each state. Given the state $(q,r,a,t)$, based on the constrained solution, an active action (a packet is transmitted) is taken if $q > u^*(r,a,t)$ and the action is passive (no transmission) otherwise. Define $r_{\max}$ as the maximum allowed transmission rate.

*Theorem VI.2:* As $\beta \to 0$, the solution $u^*(r,a,t) \to 0$ and $u^*(r,a,t) = \infty$ for $\beta > \frac{\alpha r_{\max}}{1-\alpha}$.

*Proof:* (Sketch) As $\beta \to 0$, Equation 4 implies that the cost of serving decreases to zero except that the constraint should be satisfied. Thus the solution would be to serve as much as possible, i.e., $s(x) \to \min(q, \lfloor tr \rfloor)$. Thus the action is active in any state where it is possible to do so. To show the other part, it is enough to show that $W_n(q,r,a,t) \leq \frac{1}{1-\alpha}$. Since $W_0(q,r,a,t) = 0$, if $\beta > \frac{\alpha}{1-\alpha} r_{\max}$, then $u_0^*(r,a,t) = \infty$

and $W_1(q,r,a,t)=1$. Let $W_n(q,r,a,t) \le \frac{1}{1-\alpha}$. Then $u_n^*(r,a,t)=\infty$ and $W_{n+1}(q,r,a,t) \le 1+\frac{\alpha}{1-\alpha}$. By induction hypothesis it follows that $W(q,r,a,t) \le \frac{1}{1-\alpha}$ and $u^*(r,a,t)=\infty$. Thus all actions are passive. ■

Given a state $x=(q,r,a,t)$ with $q>0$, the amount served $s(x)$ decreases to zero as $\beta$ increases and $s(x)=0$ for $\beta > \frac{\alpha}{1-\alpha} r_{\max}$. This is natural to expect since the larger is the $\beta$, the higher is penalty for transmitting.

*Theorem VI.3:* If $\beta < \frac{\alpha r_{\max}}{1-\alpha}$, then the solution $u^*(r,a,t)=0$ for $r=r_{\max}$.

*Proof:* (Sketch) Observe that for $n$ (the iteration index) satisfying $\frac{1-\alpha^n}{1-\alpha} < \frac{\beta}{\alpha r_{\max}}$, the optimal policy $u_n^*(r,a,t)=\infty$ and $W_n(q,r,a,t)=\frac{1-\alpha^n}{1-\alpha}$. Since $\beta < \frac{\alpha r_{\max}}{1-\alpha}$, $k=\min\{n : \frac{1-\alpha^n}{1-\alpha} \ge \frac{\beta}{\alpha r_{\max}}\}$ is finite. It follows that $u_k^*(r_{\max},a,t)=0$ and $W_{k+1}(q,r,a,t) \ge 1+\frac{\beta}{r_{\max}}$. Since $W_n(\cdot)$ is increasing in $n$, it can be shown that for $\beta < \frac{\alpha r_{\max}}{1-\alpha}$, $W_n(q,r,a,t) \ge 1+\frac{\beta}{r_{\max}}$ for all $n>k$. This would imply that $u_n^*(r_{\max},a,t)=0$ for all $n>k$. Hence the results follows by induction. ■

*Lemma VI.1:* $W_n(q,r,a,t)$ is nondecreasing in $q$ for each $n$.

*Proof:* The result follows from the convexity of $V_n(q,r,a,t)$ in $q$. ■

*Theorem VI.4:* The unconstrained minimizer $u^*(r,a,t)$ is monotonically nondecreasing with $\beta$.

*Proof:* We introduce the parameter $\beta$ as a variable in the functions defined earlier. Observe that the recursive algorithm stated for $W_n(q,r,a,t)$ in the previous section is equivalent to the following recursion (obtained by dividing throughout by $\beta$ as $\beta>0$). Initialize $W_0(q,r,a,t,\beta)=0$. Let $u_n^*(r,a,t,\beta)$ be the value of $u$ that solves the following inequalities,

$$\alpha E_{a,r,t} W_n(u+A,R,A,T,\beta) \le \frac{1}{r} \le \alpha E_{a,r,t} W_n(u+1+A,R,A,T,\beta). \tag{7}$$

Furthermore,

- If $q \le u_n^*(r,a,t,\beta)$, $W_{n+1}(q,r,a,t,\beta)=\frac{1}{\beta}+\alpha E_{a,r,t} W_n(q+A,R,A,T,\beta)$.
- If $u_n^*(r,a,t,\beta) < q \le \lfloor tr \rfloor + u_n^*(r,a,t,\beta)$, $W_{n+1}(q,r,a,t,\beta)=\frac{1}{\beta}+\frac{1}{r}$.
- If $q > u_n^*(r,a,t,\beta)+\lfloor tr \rfloor$, $W_{n+1}(q,r,a,t,\beta)=\frac{1}{\beta}+\alpha E_{a,r,t} W_n(q-\lfloor tr \rfloor+A,R,A,T,\beta)$.

Using Lemma VI.1, it follows from (7) that in order to show that $u_n^*(r,a,t,\beta)$ is monotonically nondecreasing in $\beta$, it is enough to show that the function $W_n(q,r,a,t,\beta)$ is nonincreasing in $\beta$ for all $n$. We show this by induction. The function $W_0(u,r,a,t,\beta)=0$. Let $W_n(q,r,a,t,\beta)$ be nonincreasing in $\beta$. This implies $E_{a,r,t} W_n(q+A,R,A,T,\beta)$ is nonincreasing in $\beta$ and $u_n^*(r,a,t,\beta)$ is monotone nondecreasing in $\beta$. Now, given $(q,r,a,t)$, the above recursion seen as a function of $\beta$ is,

- For $\beta$ where $u_n^*(r,a,t,\beta)+\lfloor tr \rfloor < q$, $W_{n+1}(q,r,a,t,\beta)=\frac{1}{\beta}+\alpha E_{a,r,t} W_n(q+A-\lfloor tr \rfloor,R,A,T,\beta)$.
- For $\beta$ where $u_n^*(r,a,t,\beta) < q \le \lfloor tr \rfloor + u_n^*(r,a,t,\beta)$, $W_{n+1}(q,r,a,t,\beta)=\frac{1}{\beta}+\frac{1}{r}$.
- For $\beta$ where $u_n^*(r,a,t,\beta) \ge q$, $W_{n+1}(q,r,a,t,\beta)=\frac{1}{\beta}+\alpha E_{a,r,t} W_n(q+A,R,A,T,\beta)$.

It follows from the definition of the minimizer and (7) that for the domain of $\beta$ where the first item holds, $\alpha E_{a,r,t} W_n(q+A-tr,R,A,T,\beta) \ge \frac{1}{r}$ and for the domain of $\beta$ where the third item holds $\alpha E_{a,r,t} W_n(q+$

$A, R, A, T, \beta) \leq \frac{1}{r}$. Thus combining this with the hypothesis that $E_{a,r,t} W_n(q + A, R, A, T, \beta)$ is nonincreasing in $\beta$ implies that $W_{n+1}(q, r, a, t, \beta)$ is nonincreasing in $\beta$ and the result follows. ■

From Theorems VI.2 and VI.4 we obtain the following conclusion:

*Corollary VI.1:* The system is indexable. ■

Given a state $(q, r, a, t)$, define the index $\nu(q, r, a, t)$ as the largest value of $\beta$ for which $u^*(r, a, t, \beta) < q$. It is essentially that value of $\beta$ where a transition is made from an active action to a passive action in the state $(q, r, a, t)$. It follows from Theorems VI.2 and VI.3 that for $r = r_{\max}$, $\nu(q, r, a, t) = \frac{\alpha r_{\max}}{1-\alpha}$. Note that the index is independent of the queue lengths when $r = r_{\max}$.

*Lemma VI.2:* The index associated with the state $(q, r, a, t)$ when the weight is $\omega$, is $\nu(q, r, a, t, \omega) = \omega\nu(q, r, a, t)$. ■

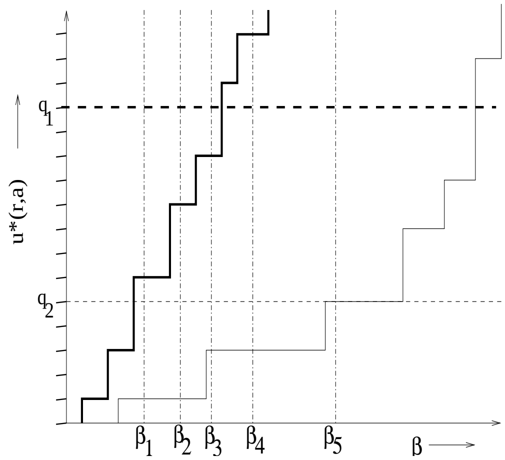
### *E. Index Based Heuristic Policy*

The transition probability matrices associated with device $i$ are $P_i^{(r)}$ and $P_i^{(a)}$. Let $\nu_i(q_i, r_i, a_i, t, \omega_i)$ be the index for device $i$ when it is in state $(q_i, r_i, a_i, t)$ and the weight is $\omega_i$. Let $u_i^*(r_i, a_i, t, \beta)$ be the solution in that state for the relaxed problem. Given the state of the system $(\mathbf{q}, \mathbf{r}, \mathbf{a}, t)$, the controller has to decide upon who should send and how much in a frame of duration $t$ seconds. Select a value for $\beta$. The amount of data served from user $i$ is $s_i(q_i, r_i, a_i, t, \beta)$. The time taken to transmit this data is $\sum_{i \in N_S} \frac{s_i(q_i, r_i, a_i, t, \beta)}{r_i}$. This could exceed the frame boundary or fall short of it depending on the choice of $\beta$. We know from Indexability that for $\beta$ arbitrary large, the solution $u_i^*(\cdot)$ is infinite and thus $s_i(\cdot)$ is zero implying that the frame time is zero. While for $\beta \to 0$, $s_i(\cdot) \to \min(q_i, \lfloor tr_i \rfloor)$, the frame boundary could be exceeded depending on the choice of $q_i$. Since as $\beta$ decreases, $s_i(q_i, r_i, a_i, t, \beta)$ increases and thus the frame time utilized increases. Thus the controller has to tune $\beta$ such that the available frame time is maximally utilized or the frame boundary constraint is met. An example is given in figure 6. Note that $s_i(q_i, r_i, a_i, t, \beta)$ has only one degree of freedom because fixing $\beta$ fixes $s_i(\cdot)$ for all $i$.

The tuning of $\beta$ is in general not an easy task. But since $u^*(r, a, t, \beta)$ is monotone nondecreasing in $\beta$, we have a simpler form for the policy.

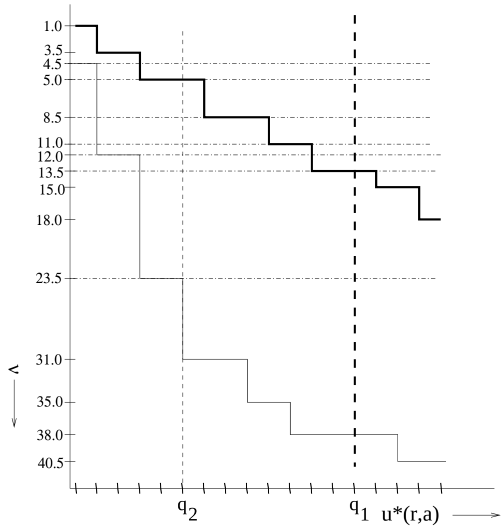
**Index Policy:** *Given the state $(\mathbf{q}, \mathbf{r}, \mathbf{a}, t)$, a user with the largest value of $\nu_i(q_i - 1, r_i, a_i, t, \omega_i)$ transmits one packet. Let $j = \arg\max_i \nu_i(q_i - 1, r_i, a_i, t, \omega_i)$. The state changes to $(\mathbf{q} - e_j, \mathbf{r}, \mathbf{a}, t)$, where $e_j$ is the unit vector with one at the $j^{th}$ entry and rest are all zero. This continues till the frame boundary is exceeded or there is no data in the buffers. The ties are broken probabilistically. The procedure is shown in Figure 7 for the example considered earlier and shown in Figure 6.*

**Remark:** Consider a case where the rate available for transmission is fixed but it can be different for different devices. Let $r_i$ be the transmission rate for device $i$. The index policy obtained above will order the




| $\beta$ | $s_1$ | $s_2$ | $\eta$ |
|---|---|---|---|
| $\beta_5$ | 0 | 0 | 0 |
| $\beta_4$ | 0 | 2 | 0.25 |
| $\beta_3$ | 2 | 2 | 0.5 |
| $\beta_2$ | 4 | 4 | 1.0 |
| $\beta_1$ | 7 | 4 | 1.375 |

Fig. 6. Consider two devices with state $(\mathbf{q}, \mathbf{r}, \mathbf{a}, t)$ with $q_1 = 12$ and $q_2 = 5$ as shown in the figure. Let the transmission times be the same for each packet. Suppose that a maximum of eight packets can be transmitted in the frame. The darker staircase function represents $u^*(\cdot)$ for device 1 while the other staircase corresponds to that of device 2. The table shows the optimal choice of $s_1$ and $s_2$, the number of packets that are sent in the frame from the two devices for various choices of $\beta$. The variable $\eta$ represents the fraction of frame time utilized. For $\beta > \beta_5$, it is optimal to serve nothing whereas $\beta = \beta_1$ the frame constraint is violated as $\eta > 1$. Thus we operate at $\beta = \beta_2$ where $s_1 = s_2 = 4$ and the frame boundary is also met.




| $q_1$ | $q_2$ | $\nu_1$ | $\nu_2$ | $s_1$ | $s_2$ | $\eta$ |
|---|---|---|---|---|---|---|
| 12 | 5 | 13.5 | 23.5 | 0 | 1 | .125 |
| 12 | 4 | 13.5 | 23.5 | 0 | 1 | .25 |
| 12 | 3 | 13.5 | 12.0 | 1 | 0 | .375 |
| 11 | 3 | 13.5 | 12.0 | 1 | 0 | .5 |
| 10 | 3 | 11.0 | 12.0 | 0 | 1 | .625 |
| 10 | 2 | 11.0 | 12.0 | 0 | 1 | .75 |
| 10 | 1 | 11.0 | 4.5 | 1 | 0 | .875 |
| 9 | 1 | 11.0 | 4.5 | 1 | 0 | 1.0 |
| 8 | 1 | 8.5 | 4.5 | 0 | 0 | – |

Fig. 7. The figure is a flipped version of figure 6 The table shows the index values $\nu_1$ and $\nu_2$ for the two users as the function of their queue lengths calculated from the figure as per the definition of indices. The one with the larger index send one packet and the queue length changes. The whole procedure as described earlier is shown as a table. The algorithm stops when $\eta = 1$, $q_1 = 8$ and $q_2 = 1$. The procedure shown in Figure 6 is equivalent to the one shown in this figure.

transmissions in decreasing order of $\omega_i r_i$ and the one with the highest order transmits till it finishes or the frame boundary is exceeded. Note that this is identical to the well known $c\mu$-rule [12].

*It is easy to verify the conditions for the existence of a stationary average cost optimal policy* $\{S[k]\}$ *(refer [11]). Further, the conditions also imply that the average optimal policy is a limit of discount optimal policies. Thus the average cost optimal policy also possess the structural properties of discount optimal policies. The number of packets transmitted in a slot is nonincreasing in* $\beta$*. Thus we have indexability and* *the indices, as defined for the discounted cost problem, defines an index policy for the average cost (mean delay) problem.*

### *F. Numerical Results*

Let us assume that there are no voice calls. The discount factor is set to $\alpha = 0.99$ implying that the long term evolution of the queue length process contribute significantly towards the performance measure. The other parameters for the numerical computation of the policy are: the frame time $T = 10$ms, the transmission rate set $\{10, 3.3, 2.5\}$ kbps. We consider two transition probability matrices for the rate process:

$$P_1 = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.99 & 0.01 & 0 \\ 0 & 0.99 & 0.01 \end{bmatrix}; \; P_2 = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.01 & 0 & 0.99 \\ 0.01 & 0.99 & 0 \end{bmatrix}$$

For the rate process governed by $P_1$, with a very large probability the rate increases from one of the lower rates to the next higher rate and then goes to one of the lower rates with equal probability whereas for the rate process governed by $P_2$, the rate process switches between the two lower rate states with high probability. Thus $P_2$ resembles a device operating far away from the AP and restricted mobility where as $P_1$ resembles a device that is highly mobile. The packet arrival process is assumed to independent and identically distributed, on-off $\{0, 40\}$ with probability $\{.5, .5\}$. Since the arrival process is i.i.d. and the frame time available is fixed to $T$ (no voice calls), the policy $u^*(r, a, t, \beta)$ is independent of $a$ and $t$. Also $u^*(r, a, t, \beta)$ for $r = r_{\max}$ is $\frac{\alpha r_{\max}}{1-\alpha} = 9.9 \times 10^5$. Figure 8 plots $u^*$ vs $\beta$ for $r = \{3.3, 2.5\}$ kbps and the rate transition probability matrices $P_1$ and $P_2$.

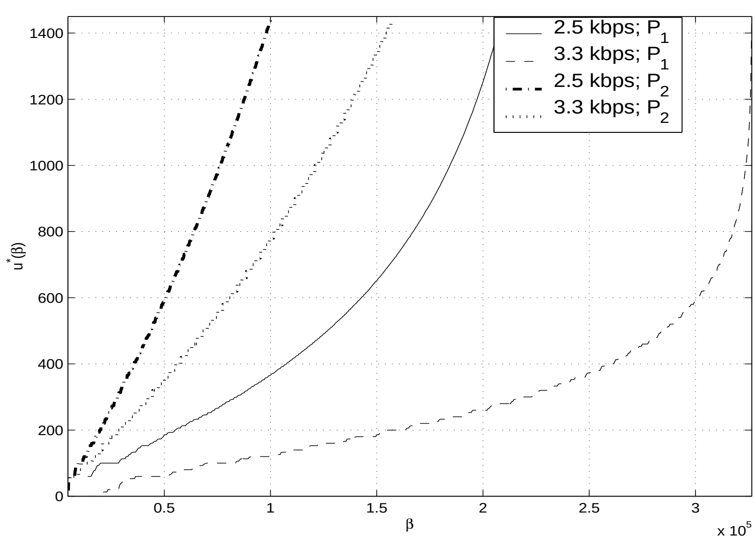



Fig. 8. Plots are used for computing indices $\nu$. For example consider two devices with the rate transition probability matrices $P_1$ and $P_2$. The weights are 1 for both the devices, $q_1 = q_2 = 600$, $r_1 = 2.5$ and $r_2 = 3.3$ kbps. The indices $\nu_1 = 14.35 \times 10^4$ and $\nu_2 = 8 \times 10^4$. This shows that device 1 has priority over 2 even when $r_2 > r_1$. If one of the device has a rate of 10 kbps, then the service effort is applied to it as much as possible since the index is the largest independent of the queue length.

For the scenario discussed above, we compared the performance of the index policy with that of a round robin policy, a weighted round robin policy that serves three packets of device 2 for each packet of device

1, a stabilizing policy $\omega_i q_i r_i$ [3]. For a fixed initial state $\mathbf{z} = (\mathbf{q}, \mathbf{r})$ with $q_1 = q_2 = 0$ and $r_1 = r_2 = 2.5$ kbps, the costs $(1-\alpha)V_\alpha(\mathbf{z})$ are $107$, $398$, $327$ and $128$ respectively. Define $T_S[k]$ as the time taken by the streaming traffic during the $k^{th}$ frame when using the index policy.

## VII. Analysis: File Transfers

The subframe time $T[k]$ available for the file transfer sessions during $k^{th}$ frame is $\tau - T_S[k] - T_V[k]$. Note that $T[k] > \tau_T[k]$ and the process $T[k]$ is a finite state Markov chain. Let $\mathbf{\Gamma}$ be the transition probability matrix for $T[k]$. A realization of the random variable $T[k]$ is denoted by $t$. The system model was discussed in Section IV. First, we look at the problem where the data to be served is fluid rather than packet or that the packets can be arbitrarily fragmented. Also assume that only one of the devices can transmit during the subframe; since the queues are always backlogged no frame time is wasted. We will later use the results obtained for the fluid model to provide index policies for the packet model discussed in the Section IV. In these packets service policies more than one device would be able to transmit in a frame.

If a device transmits in a particular frame, the AP learns about that user channel state, or equivalently the transmission rate; otherwise the information available at the AP is old information from when the device last transmitted. Thus this is a case of a system with partial state information. Let $P = P^{(r)}$ be the transition probability matrix for the rate process. If $r_i$ is the rate at which device $i$ last transmitted, and $d_i$ is the number of frames since the last transmission, the AP has the information about the probability measure on the rate space for channel to/from device $i$ in any frame. The measure is $\pi_i(r) = P^{d_i}_{r_i}(r)$, a row corresponding to rate $r_i$ of the matrix $P^{d_i}$. The state of the system is represented by $\mathbf{x} = (\mathbf{r}, t, \mathbf{d})$ where $t$ is the subframe time.

Let the action in frame $k$ be $S_i[k]$ where $S_i[k] \in \{0, 1\}$. If $S_i[k] = 1$, the reward earned is the amount of fluid released $Z_i(\mathbf{x}) = \omega_i \sum_r rtP^{d_i}_{r_i}(r)$ while no reward is earned for an action $S_i[k] = 0$. The summation above is over the rate set. To show the dependence of reward on the state, action and user, we use the notation $Z_i(x_i, S_i)$ to represent the reward earned for user $i$ when its state $x_i = (r_i, t, d_i)$ and an action $S_i$ is taken. The total reward is thus the sum of individual rewards. Also $\sum_{i \in N_T} S_i[k] \leq 1$ for all $k$, since only one connection is scheduled to transmit in each frame. Let $\Pi$ be the space of all Markovian policies mapping the system state to the action set $\{0, 1\}^{N_T}$. Let $\Pi^c$ be a subset of $\Pi$ that satisfies the above said constraint that at the most one user can transmit in any frame.

The problem consists of finding a scheduling policy $\pi \in \Pi^c$ that maximizes the long run time average reward rate, $Z^*(1) = \max_{\pi \in \Pi^c} \liminf_{n \to \infty} \frac{1}{n} E_\pi \left[ \sum_{k=0}^{n} \sum_{i \in N_T} Z(X_i[k], S_i[k]) \right]$ or the long run discounted reward earned, $Z^*(\alpha) = \max_{\pi \in \Pi^c} E_\pi \left[ \sum_{k=0}^{\infty} \alpha^k \sum_{i \in N_T} Z(X_i[k], S_i[k]) \right]$.

We use Whittle's relaxation and demand that at the most one user can transmit on the average. Thus the optimal value for the relaxed problem is an upper bound for the original problem's optimal value. We can now decouple the above said problem and solve it for each device. Dropping the connection index $i$, for the decoupled problem the system state is a triplet $(r,t,d)$ where $r$ is the rate at which the last transmission was made for this connection and $d$ represents the time slots that have elapsed since the last transmission for this connection and $t$ is the time available in the current frame. Let $\nu$ be the Lagrange multiplier associated with the relaxed constraint, representing the reward offered for not transmitting. Without loss of generality, take $\omega = 1$. The discounted cost objective is to obtain a sequence $S[k]$ that maximizes, $E\left[\sum_{k=0}^{\infty}\alpha^k(Z(X[k],S[k])-\nu S[k])\right]$ or equivalently, $E\left[\sum_{k=0}^{\infty}\alpha^k\left(\sum_r\left(rT[k]S[k]\;P_{r[k]}^{d[k]}(r)\right)-\nu S[k]\right)\right]$.

Define $V(r,t,d)$ as the optimal expected discounted reward earned when the initial state is $(r,t,d)$. Let $\mathbf{V}(d)$ be the matrix such that the entry corresponding to the $r^{th}$ row and $t^{th}$ column is $V(r,t,d)$. Then the expected reward with respect to the variable $t$, $E_t[V(r,T,d)]$, is a element in the $r^{th}$ row and the $t^{th}$ column of $\mathbf{V}(d)\mathbf{\Gamma}'$ where $C'$ denote the transpose of a matrix $C$. Then the expected reward with respect to the variable $r$, $E_r[V(R,t,d)]$, is the $r^{th}$ row and $t^{th}$ column of $\mathbf{PV}(d)$. Also $[C]_{i,j}$ represents the $i^{th}$ row and $j^{th}$ column of the matrix $C$.

Define a matrix $\mathbf{M}$ with rows representing rate $r$ and column representing subframe time $t$ and set $[\mathbf{M}]_{r,t} = rt$. Let $\mathbf{1}$ be the matrix with all entries equal to one. The discounted cost optimality equation for the said relaxed problem is, $V(r,t,d) = \max\{[\mathbf{P}^d(\mathbf{M}-\nu\mathbf{1}+\alpha\mathbf{V}(1)\mathbf{\Gamma}')]_{r,t}, \alpha[\mathbf{V}(d+1)\mathbf{\Gamma}']_{r,t}\}$. If we define that the maximization is taken component wise, we can rewrite the above equation in a more compact form as,

$$\mathbf{V}(d) = \max\{\mathbf{P}^d(\mathbf{M}-\nu\mathbf{1}+\alpha\mathbf{V}(1)\mathbf{\Gamma}'), \alpha\mathbf{V}(d+1)\mathbf{\Gamma}'\}. \tag{8}$$

Observe from the above equation that $\mathbf{V}(d)$ is given in terms of $\mathbf{V}(d+1)$. Thus we can expand the right hand side of the above equation and get

$$\mathbf{V}(d) = \max_{k\geq d}\{\alpha^{k-d}\mathbf{P}^k(\mathbf{M}-\nu\mathbf{1}+\alpha\mathbf{V}(1)\mathbf{\Gamma}')\mathbf{\Gamma}'^{k-d}\}. \tag{9}$$

Note that if we know $\mathbf{V}(1)$ all others can be easily determined and hence so can the solution. Thus, the objective is to first determine $\mathbf{V}(1)$.

$$\mathbf{V}(1) = \max_{k\geq 1}\{\alpha^{k-1}\mathbf{P}^k(\mathbf{M}-\nu\mathbf{1}+\alpha\mathbf{V}(1)\mathbf{\Gamma}')\mathbf{\Gamma}'^{k-1}\}. \tag{10}$$

Consider the corresponding discounted cost value iteration algorithm for evaluating $\mathbf{V}(1)$.

$$\mathbf{V}_n(1) = \max_{k\geq 1}\{\alpha^{k-1}\mathbf{P}^k(\mathbf{M}-\nu\mathbf{1}+\alpha\mathbf{V}_{n-1}(1)\mathbf{\Gamma}')\mathbf{\Gamma}'^{k-1}\}. \tag{11}$$

with $\mathbf{V}_0(1) = \mathbf{0}$, the zero matrix. It is well known that the $\mathbf{V}_n(1)$ converges to $\mathbf{V}(1)$.

Recalling the Lagrange multiplier $\nu$, note that a large value of $\nu$ discourages transmissions (i.e., encourages passivity). Let us associate a value $\nu(r,t,d)$ with state $(r,t,d)$ representing the value of making a transmission attempt when the state is $(r,t,d)$. The device with the highest such value will be polled for transmission. Since the channel is reciprocal, the device would estimate the channel on the polled frame and transmit at the estimated rate. The AP would also come to know about the rate as the data transmission starts from the device. The value $\nu(r,t,d)$ is that choice of $\nu$ for which the optimal action in state $(r,t,d)$ makes a transition from active to passive, i.e., the maximizer in Equation 9 changes from $k^* = d$ to some number larger than $d$. This can be seen as that value of $\nu$ which makes the choice of $k^* = d$ and $k^* > d$ equally attractive. In order to carry this out, we need to show indexability ( Definition VI.1).

*Theorem VII.1:* If $\nu > \max \mathbf{M}$ then all the states are passive.

*Proof:* The hypothesis implies that $\mathbf{M} - \nu\mathbf{1} < 0$. Thus if $\mathbf{V}_0(1) = 0$, then Equation 11 implies that $\mathbf{V}_1(1) = 0$ and the maximizer is $k^* = \infty$. Thus by induction it would follow that $\mathbf{V}(1) = 0$ and the maximizer is $k^* = \infty$. Thus all the states are passive. ■

*Theorem VII.2:* The optimal value function $V(r,t,d)$ is convex nonincreasing in $\nu$.

*Proof:* Owing to the representation in Equation 9, it is enough to show the statement for the case $d = 1$. We show that $V(r,t,1)$ has the said property by induction. In the matrix notation each function needs to be shown to have the desired property. We know that the convex combination of convex nonincreasing functions is convex nonincreasing. Since $\mathbf{V}_0(1) = \mathbf{0}$, the statement holds. Let $\mathbf{V}_n(1)$ have the said property. Consider Equation 11. Note that each component of the matrix within the braces is convex nonincreasing in $\nu$ for each $k$. As $\mathbf{V}_n(1)$ is maximum over such functions, $\mathbf{V}_n(1)$ is also convex and nonincreasing in $\nu$. Thus by induction hypothesis, $\mathbf{V}(\mathbf{1})$ has the said property. ■

*Theorem VII.3:* The indices $\nu(r,t,d) \geq [\mathbf{P}^d\mathbf{M}]_{r,t}$.

**Remark:** Note that $[\mathbf{P}^d\mathbf{M}]_{r,t}$ is the expected value of $R[d]$ given that the system starts in state $r$ at time $0$ multiplied by the frame time $t$.

*Proof:* If we show that in Equation 8, $\mathbf{P}^d\mathbf{V}(1) \geq \mathbf{V}(d+1)$, then we are done since that would imply that $\mathbf{P}^d\mathbf{V}(1)\mathbf{\Gamma}' \geq \mathbf{V}(d+1)\mathbf{\Gamma}'$. Thus all the states are active for $\mathbf{P}^d\mathbf{M} > \nu\mathbf{1}$. Hence $\nu(r,t,d)$ should be greater than or equal to $[\mathbf{P}^d\mathbf{M}]_{r,t}$. We have,

$$\begin{aligned}
\mathbf{V}(d+1) &= \max_{k\geq(d+1)}\{\alpha^{k-d-1}\mathbf{P}^k(\mathbf{M}-\nu\mathbf{1}+\alpha\mathbf{V}(1)\mathbf{\Gamma}')\mathbf{\Gamma}'^{k-d-1}\}, \\
&= \max_{k\geq(d+1)}\{\mathbf{P}^d\alpha^{k-d-1}\mathbf{P}^{k-d}(\mathbf{M}-\nu\mathbf{1}+\alpha\mathbf{V}(1)\mathbf{\Gamma}')\mathbf{\Gamma}'^{k-d-1}\}, \\
&= \max_{k\geq 1}\{\mathbf{P}^d\alpha^{k-1}\mathbf{P}^k(\mathbf{M}-\nu\mathbf{1}+\alpha\mathbf{V}(1)\mathbf{\Gamma}')\mathbf{\Gamma}'^{k-1}\} \quad \leq \quad \mathbf{P}^d\mathbf{V}(1).
\end{aligned}$$

where the last inequality follow from Equation 10 and Jensen's inequality. ■

The above results provide upper and lower bounds on the index value. Next we ask the question whether the system is indexable, i.e, is it true that once a state $(r,t,d)$ that has been made passive at say $\nu(r,t,d) = \nu_0$, it cannot be made active by increasing $\nu > \nu_0$. In following example we show that even for the case where the process $T[k]$ is constant, it is a difficult question to answer.

Let $T[k]$ be constant, say, normalised to $1$. The value function $\mathbf{V}(1)$ will now be a vector. Let $R$ be the vector of all possible transmission rates. The optimality equation is

$$\mathbf{V}(d) = \max\{\mathbf{P}^d(\mathbf{R} - \nu\mathbf{1} + \alpha\mathbf{V}(1)), \alpha\mathbf{V}(d+1)\} = \max_{k \geq d}\{\alpha^{k-d}\mathbf{P}^k(\mathbf{R} - \nu\mathbf{1} + \alpha\mathbf{V}(1))\}. \tag{12}$$

Given a vector of integers say $\mathbf{n}$. Let $\mathbf{A}$ be a square matrix. Define $\mathbf{A}^{\mathbf{n}}$ as a matrix whose $i^{th}$ row is the $i^{th}$ row of the matrix $\mathbf{A}^{n_i}$. Equation 12 for $\mathbf{k} = \mathbf{n}$ and $d = 1$ can be written as $\mathbf{V}(\mathbf{1}) = \frac{1}{\alpha}(\mathbf{I} - (\alpha\mathbf{P})^{\mathbf{n}})^{-1}(\alpha\mathbf{P})^{\mathbf{n}}(\mathbf{R} - \nu\mathbf{1})$. Thus, $\mathbf{V}(\mathbf{1}) = \max_{\mathbf{n} \geq \mathbf{1}}\left\{\frac{1}{\alpha}(\mathbf{I} - (\alpha\mathbf{P})^{\mathbf{n}})^{-1}(\mathbf{R} - \nu\mathbf{1})\right\} - \frac{1}{\alpha}(\mathbf{R} - \nu\mathbf{1})$.
Let $\mathbf{n_1}, \mathbf{n_2}$ be optimal values of $\mathbf{n}$ for $\nu_1, \nu_2$ respectively with $\nu_1 < \nu_2$. Then

$$\frac{1}{\alpha}(\mathbf{I} - (\alpha\mathbf{P})^{\mathbf{n_1}})^{-1}(\mathbf{R} - \nu_1\mathbf{1}) - \frac{1}{\alpha}(\mathbf{R} - \nu_1\mathbf{1}) \geq \frac{1}{\alpha}(\mathbf{I} - (\alpha\mathbf{P})^{\mathbf{n_2}})^{-1}(\mathbf{R} - \nu_1\mathbf{1}) - \frac{1}{\alpha}(\mathbf{R} - \nu_1\mathbf{1}),$$
$$\frac{1}{\alpha}(\mathbf{I} - (\alpha\mathbf{P})^{\mathbf{n_2}})^{-1}(\mathbf{R} - \nu_2\mathbf{1}) - \frac{1}{\alpha}(\mathbf{R} - \nu_2\mathbf{1}) \geq \frac{1}{\alpha}(\mathbf{I} - (\alpha\mathbf{P})^{\mathbf{n_1}})^{-1}(\mathbf{R} - \nu_2\mathbf{1}) - \frac{1}{\alpha}(\mathbf{R} - \nu_2\mathbf{1}).$$

Adding the above equations we get, $(\mathbf{I} - (\alpha\mathbf{P})^{\mathbf{n_1}})^{-1}(\nu_2 - \nu_1)\mathbf{1} \geq (\mathbf{I} - (\alpha\mathbf{P})^{\mathbf{n_2}})^{-1}(\nu_2 - \nu_1)\mathbf{1}$. Equivalently, $(\mathbf{I} - (\alpha\mathbf{P})^{\mathbf{n_1}})^{-1}\mathbf{1} \geq (\mathbf{I} - (\alpha\mathbf{P})^{\mathbf{n_2}})^{-1}\mathbf{1}$.
We now need to show that $n_1 \leq n_2$. Unfortunately this is not true. Consider the following counterexample. Let $\alpha = 0.99, P = \{0.01, 0.99; 0.99, 0.01\}, \mathbf{n}_1 = \{2, 1\}$ and $\mathbf{n}_2 = \{4, 1\}$. Then $(I - (\alpha P)^{\mathbf{n_1}})^{-1}\mathbf{1} = \{5.2, 4.8\}$ and $(I - (\alpha P)^{\mathbf{n_2}})^{-1}\mathbf{1} = \{2.96, 3.67\}$.

Since the above condition is a sufficient condition for Indexability, the above counterexample does not imply that the system is not indexable. But it is difficult to prove or disprove the Indexability. The following definition weaken the indexability condition.

*Definition VII.1:* The system is said to be **weakly indexable** if for each system state $x$ there exists a value $\nu(x)$ such that a transition from active to passive is made at $\nu(x)$ and the optimal action in that state is passive for all $\nu > \nu(x)$. The value $\nu(x)$ defines the weak index for state $x$. ■
Note that the definition is consistent, i.e., if the system is indexable then the weak index agrees with the index. Further, weak indexability will be implied by the existence of a finite $\nu^*$ such that for all $\nu > \nu^*$, the optimal action is passive, for all the system states. Thus in view of Theorem VII.1, the fluid system with varying subframe lengths as considered earlier is weakly indexable.

### *A. Packet Model*

Now consider the actual problem, where packets need to be sent instead of fluid. There is a trade off. The polling stations can ask for only one packet per device until the subframe boundary is met. This way it could get fresh channel state information for many links. But it could result in potentially lower throughput than that available on good links since it would not efficiently utilize only those links that have a higher rate.

The system state is $(\mathbf{r}, t, \mathbf{d})$ with $r_j$ represents the number of packets that can be transmitted per unit time if the whole service effort is applied to device $j$. The schedule should decide upon $\mathbf{S}[n]$, the number of packets from each device that should be transmitted in a subframe of length $t$ units. The sequence $\{\mathbf{S}[n]\}$ should satisfy the subframe boundary constraint, i.e., $\sum_{i=1}^{N_T} \frac{S_i[n]}{R_i[n]T_i[n]} \leq 1$ for all $n$. We relax the above constraint. The approach is similar to the one carried out earlier. Given that the rate is $r$, the penalty for transmitting $s$ packets would be the fraction of subframe time used $\nu \frac{s}{rt}$, whereas the reward is the number of packets transmitted $s$. Note that $s \in \{0, 1, \cdots, \lfloor rt \rfloor\}$. Based on the analysis for the fluid model, we have the following optimality equation for the decoupled problem,

$$V(r,t,d) = \max\left\{ \sum_{r'} P^d_{r,r'} \left( \max_{1 \leq s \leq \lfloor r't \rfloor} \left\{ s - \nu \frac{s}{r't} \right\} + \alpha \sum_{t'} \Gamma_{t,t'} V(r',t',1) \right) \ , \ \alpha V(r,t,d+1) \right\}.$$

Note that the inner maximizer can either be $1$ or $\lfloor r't \rfloor$ depending on the choice of $\nu$. If $\nu > r't$ then $s = 1$, whereas $s = \lfloor r't \rfloor$ otherwise. Thus $\nu = rte$ is a crossover point. The optimality equation is,

$$V(r,t,d) = \max\left\{ \sum_{r'} P^d_{r,r'} \left( \max\left\{ \left(1 - \frac{\nu}{r't}\right), \left(\lfloor r't \rfloor - \frac{\nu \lfloor r't \rfloor}{r't}\right) \right\} + \alpha \sum_{t'} \Gamma_{t,t'} V(r',t',1) \right), \alpha V(r,t,d+1) \right\}.$$

Let us relate this equation to Equation 8. The matrix $\mathbf{M}$ in Equation 8 has entries $\mathbf{M}_{r,t} = rt$. Define another matrix $\mathbf{M}(\nu)$ such that

$$\mathbf{M}(\nu)|_{r,t} = \max\left\{ \left(1 - \frac{\nu}{rt}\right), \left(\lfloor rt \rfloor - \frac{\nu \lfloor rt \rfloor}{rt}\right) \right\}.$$

The optimality equation can now be written in a compact form (similar to the one in Equation 8) as,

$$\mathbf{V}(d) = \max\{\mathbf{P}^d(\mathbf{M}(\nu) + \alpha \mathbf{V}(1)\mathbf{\Gamma}'), \alpha \mathbf{V}(d+1)\mathbf{\Gamma}'\}. \tag{13}$$

The analysis approach is same the as that for the fluid model. On similar lines one can show that the system is weakly indexable. Let $\nu_o(r,t,d)$ be the weak indices for the above problem (Equation 13).

Then, given that a state $(r,t,d)$ is active (transmit one packet), one has to decide between transmitting only one packet or occupying the rest of the subframe ($s = 1$ or $s = \lfloor r't \rfloor$). As discussed earlier, the transition from $s = \lfloor r't \rfloor$ to $s = 1$ occurs at $\nu = r't$. Once a packet has been transmitted, the information regarding the current transmission rate (i.e., $r'$) is available at the polling station. Thus given $r'$, define an index associated with transmitting $s = \lfloor r't \rfloor$ as $\nu_a(r,t,d,r')$. Thus $\nu_a(r,t,d,r') = \min(\nu_o(r,t,d), r't)$. But we demanded that the decisions have to be made at the start of the frame, and should not make use of any information that is available subsequently during the frame. The above policy makes use of the information $r'$ that is only available after a packet has been transmitted. Thus the decisions do depend upon the state evolution during the frame. If we restrict ourself to make all decisions at the start of the frame itself, then the policy above needs to be appropriately modified. Though it would result in a loss of throughput as fresh information which is potentially available is not being used. The modified policy is $\nu_a(r,t,d) = \min(\nu_o(r,t,d), t[P^d\mathbf{R}]_r)$. This is appropriate as the best possible information available about

$r'$ at the start of the frame is the conditional expected rate conditioned on $(r, d)$. Also, along the lines of the proof of Theorem VII.3, we have $\nu_o(r,t,d) \geq t[P^d\mathbf{R}]_r)$. Thus $\nu_a(r,t,d) = t[P^d\mathbf{R}]_r)$.

Then the scheduling algorithm is as follows. Let device $j$ have weight $\omega_j$. Let the system state be $\{(r_j, t, d_j); j = \{1, 2, \cdots, N_T\}\}$. The index for device $j$ is a pair $(\omega_j \nu_a(r_j, t, d_j), \omega_j \nu_o(r_j, t, d_j))$. Stack the indices $\omega_j \nu_o(r_j, t, d_j)$ in a table. First, the one with the largest entry in this table transmits one packet. In case of a tie, the one with largest delay (absolute delay and not the number of slots) transmit a packet. Let device $k$ have the maximum entry. Replace the entry $k$ by $\omega_k \nu_a(r_k, t, d_k)$. Repeat the procedure until the subframe boundary is met. After completion of the subframe update the absolute delay values by the latest time stamp of the start of a packet transmission from each device. We need to track the absolute delays in order to break the ties. Update the rate vector $r$ for those who transmitted in the subframe. Also reset $d = 1$ for those who transmitted in the subframe whereas $d = d + 1$ for those who did not transmit in the subframe.

Consider a scenario where information regarding the available transmission rates are known at all times. The optimal policy would then be to transmit at the maximum rate available and the one who has the maximum rate transmits. The ties can be broken probabilistically or the one among the tied node that has the longest delay transmits. Let $\pi$ be the steady state probability distribution of the transmission rates available and let $R$ be the random variable representing rates. Define a random variable $\hat{R}$ equal to the maximum of $N_T$ independent random variables $R$. The average throughput per user would be the mean of $\hat{R}$. A round-robin polling strategy that does not use any state information would yield an aggregate throughput equal to the average of all the available transmission rates. We define another simple index policy called the "Conditional expected rate policy" with the indices defined as $\mu(r,t,d) = t[P^d\mathbf{R}]_r$ (the conditional expected rate given $(r,t,d)$). Note that this is same as $\nu_a(r,t,d)$. This policy has been shown to be optimal [8] in the case where the channel is modeled as being in one of the two states (good or bad), the process $T[k]$ was fixed to say $1$ and some restrictions were imposed on the choice of the transition probability matrix and the parameter $\alpha$. We provide numerical results for our index policy and compare its performance with that for the round-robin policy, the policy with perfect state information and the conditional expected rate policy.

### *B. Numerical and Simulation Results*

Let the subframe time available be fixed. Let there be three rates $\{10, 7, 4\}$ (packets per frame). Let $\alpha = 0.99$. The transition probability matrix for the rate process is $P = P_1$ as defined in the numerical example for streaming (Section VI). The plot for weak indices is shown in Figure 9. Also it was seen numerically that the system is indexable and thus the weak indices are also indices.

We consider a case where the weights $\omega_i$ are equal. Figure 10 plots the aggregate throughput versus the number of sessions for the four policies: index policy, round-robin policy, the policy based on perfect channel state information (state is known at all times) and the expected rate policy.

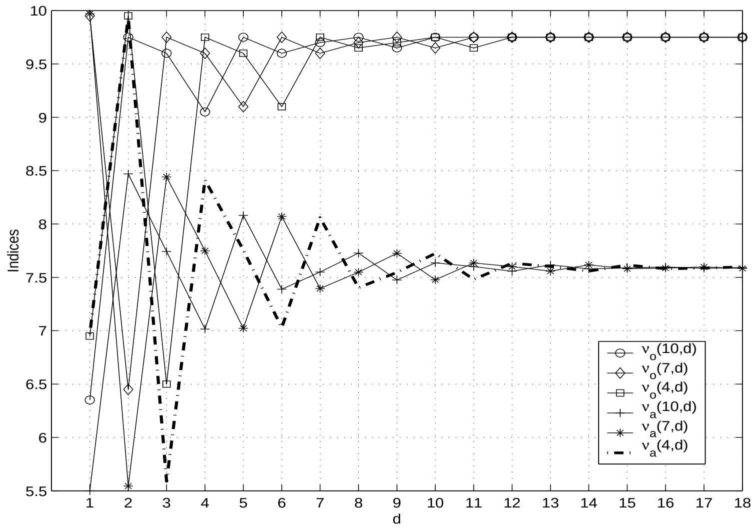



Fig. 9. The indices $\nu_o(r, d)$ and $\nu_a(r, d)$ as a function of rate $r$ and the delay $d$. For example, if the rate $r$ at which the last transmission for a connection took place is $4$ units and the number of frames since last transmission (delay) is $1$, the index values are $\nu_o(4, 1) = \nu_a(4, 1) = 6.9$. Whereas, $\nu_o(4, 3) = 6.5$ and $\nu_a(4, 3) = 5.5$.

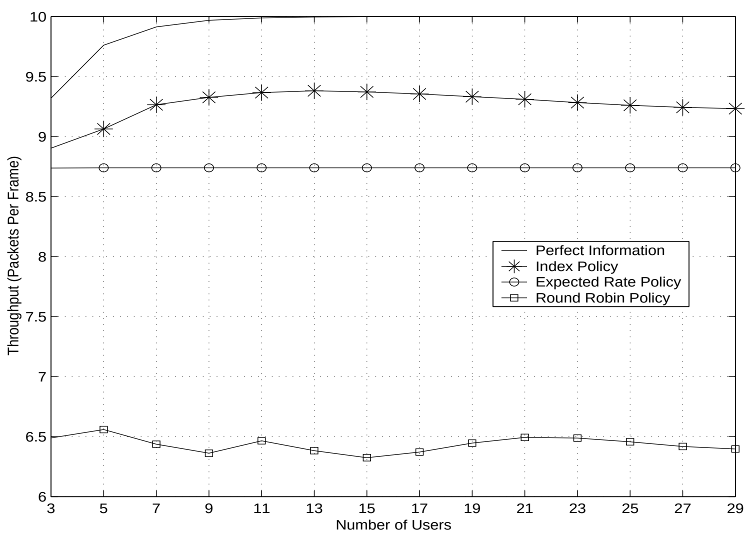



Fig. 10. Simulation results for the aggregate throughput vs the number of sessions in the network for the four policies. With $15$ active sessions in the network, the throughputs in packets per frame are: Index policy = $9.4$; Policy with perfect state information = $10$; Round robin polling = $6.4$; Conditional expected rate policy = $8.75$. Note that nearly $16\%$ of the time is wasted in case of index policy and polling policy since the packets cannot be fragmented. For the expected rate policy no time is lost. The performance of index policy would be better than that indicated if the rate set chosen has larger values. Due to delayed information and the suboptimal index policy, the throughput is $6\%$ less than the case where perfect channel knowledge is available.

## VIII. CONCLUSION

We have developed index based polling strategies for a multiaccess network over a fading wireless channel. Index policies are always desired for ease of implementation. We considered three classes of calls: voice, streaming and file transfers. An index policy is obtained in terms the system state for each of the three classes. At any time instant, the one with the highest current index transmits one packet. The performance of the index policy is compared with other known policies such as a round-robin strategy, a policy that stabilizes the system and some other intuitive policies. As part of future work we are interested in the development of algorithms for on-line computation of the indices. Further, these policies take care of call arrival and departures as they are index policies and indices do not change with the number of calls in the system. This is in fact the motivation for having index policies.

## REFERENCES

[1] *IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications* Nov. 1997, P 802.11.

[2] *SP-RFIv1.1-I01-990311, Data-Over-Cable Service Interface Specifications: Radio Frequency Interface Specification 1.1, Mar. 1999*

[3] *Andrews M., Kumaran K., Ramanan K., Stolyar A., Vijayakumar R., and Whiting P.: Scheduling in a Queueing System with Asynchronously Varying Service Rates.* Bell Labs Technical Report*, 2000.*

[4] *Berry Randall A., Gallager R. G.: Communication over Fading Channels with Delay Constraints.* IEEE Transaction on Information Theory*, vol. 48, no. 5, 1135-1149, May 2002.*

[5] *J. M. Capone, I. Stavrakakis: Delivering QoS requirements to traffic with diverse delay tolerances in a TDMA environment,* IEEE/ACM Transactions on Networking*, February 1999.*

[6] *Dantzig G. B.: Discrete Variable Extremum problems.* Operations Research*, 5, 1957, 266-277.*

[7] *Goyal Munish, Kumar Anurag, Sharma Vinod: Power constrained and Delay Optimal Policies for Scheduling Transmission over a Fading Channel.* IEEE INFOCOM*, 2003.*

[8] *Ger Koole, Z. Liu, R. Righter: Optimal transmission policies for noisy channels.* Operations Research*, 49:892-899, 2001*

[9] *W. K. Kuo, S. Kumar, C. C. Jay Kuo: Improved priority access, bandwidth allocation and traffic scheduling for DOCSIS cable networks,* IEEE Transactions on Broadcasting*, Dec. 2003.*

[10] *Jose Nino-Mora: Restless bandits, partial conservation laws and indexability.* Advances in Applied Probability*, 33(1), 2001, 76-98.*

[11] *Schal M.: Average Optimality in Dynamic Programming with General State Space.* Mathematics of Operations Research*, 18(1), 1993, 163-172.*

[12] *Walrand J.: An Introduction to Queueing Networks. Prentice-Hall, New Jersey, 1988.*

[13] *Weber R. R. and Weiss G.: On an index policy for restless bandits.* J. Appl. Prob.*, 27, 1990, 637-648.*

[14] *Whittle Peter: Restless bandits: Activity allocation in a changing world. In* A Celebration of Applied Probability*, J. Gani (Ed.),* J. Appl. Prob.*, 25A, 1988, 287-298.*

[15] *Yeh E. M. and Cohen Aaron S.: Throughput and Delay Optimal Resource Allocation in Multiaccess Fading Channels.* Proc. of ISIT*, Yokohama, Japan, 2003.*

## IX. Appendix

*Proof of Theorem VI.1.* Since $H(q,r,a,t)$ is a convex combination of $V(q+a,r,a,t)$, it suffices to show that $V(q,r,a,t)$ is convex in $q$. Consider the value iteration algorithm (5). For $n=0$, $V_0(q,r,a,t)=0$ hence convex. Assume $V_{n-1}(q,r,a,t)$ is convex in $q$. Fix $q$. Let $u_1$ and $u_2$ be the optimal policy for $q-1$ and $q+1$.

$$\begin{aligned}
&V_n(q+1,r,a,t)+V_n(q-1,r,a,t)\\
&= \quad 2q(1+\frac{\beta}{r})-\frac{\beta}{r}(u_1+u_2)+\alpha E_{a,r,t}[V_{n-1}(u_1+A,R,A,T)+V_{n-1}(u_2+A,R,A,T)],\\
&\geq \quad 2q(1+\tfrac{\beta}{r})-\tfrac{\beta}{r}(u_1+u_2)+\alpha E_{a,r,t}V_{n-1}(\lfloor\tfrac{u_1+u_2}{2}\rfloor+A,R,A,T)+\alpha E_{a,r,t}V_{n-1}(\lceil\tfrac{u_1+u_2}{2}\rceil+A,R,A,T),\\
&\geq^* \quad 2V_n(q,r,a,t)
\end{aligned}$$

where the inequality $(*)$ follows from the fact that the policies $\lfloor\frac{u_1+u_2}{2}\rfloor$ and $\lceil\frac{u_1+u_2}{2}\rceil$ are feasible for the state $(q,r,a,t)$. That the functions are nondecreasing can also be proved along similar lines.