# Index policies for real-time multicast scheduling for wireless broadcast systems

Vivek Raghunathan, Vivek Borkar, Min Cao and P. R. Kumar

***Abstract*—Motivated by the increasing usage of wireless broadcast networks for multicast real-time applications like video, this paper considers a canonical real-time multicast scheduling problem for a wireless broadcast LAN. A wireless access point (AP) has $N$ latency-sensitive flows, each associated with a deadline and a multicast group of receivers that desire to receive all the packets successfully by their corresponding deadlines. We consider periodic and one-shot models of real-time arrivals. The channel from the AP to each receiver is a wireless erasure channel, independent across users and slots. We wish to find a communication strategy that minimizes the total deadlines missed across all receivers, where a receiver counts a miss if it does not receive a packet by its deadline. We cast this problem as a restless bandit in stochastic control. We use Whittle's relaxation framework for restless bandits to establish Whittle-indexability for multicast real-time scheduling under the assumption of complete feedback from all receivers in every slot. For the Whittle relaxation, we show that for each flow, the AP's decision between transmitting in a slot and idling has a threshold structure. For the homogeneous case where the erasure channel to each receiver is identically distributed with parameter $p$, the Whittle index of a flow is $x_i(1-p)$ , where $x_i$ is the number of receivers who have yet to receive the current packet of flow $i$. For the general heterogeneous case in which the erasure channel to receiver $j$ has loss probability $p_j$, the Whittle index corresponding to each flow is $\sum_j (1-p_j)$, where the sum is over all multicast receivers who are yet to receive the packet. We bound the performance of the optimal Whittle relaxation with respect to the optimal wireless multicast real-time scheduler. The heuristic index policy that schedules the flow with the maximum Whittle index in each slot is simple. To relax the complete feedback assumption, we design a scalable mechanism based on statistical estimation theory that obtains the required feedback from all the receivers using a single ACK per packet transmission. The resultant policy is amenable to low-complexity implementation.***

## I. Introduction

Wireless broadcast networks are increasingly used for multicast applications like video with real-time constraints. In this paper, we study the problem of "broadcast-supporting" real-time scheduling in unreliable wireless environments in the specific context of multicast scheduling over a broadcast wireless LAN. We consider a setting in which a wireless AP is communicating information to a set of wireless stations. The applications are multicast, each flow associated with a multicast receiver group. The applications are soft real-time; each packet having a deadline by which it must be received successfully to have any utility. The wireless network is broadcast, and an AP transmission can potentially be decoded by multiple stations successfully, provided that the wireless link conditions from the AP to each such station are "good".

The AP has $N$ flows, each associated with a deadline. We study two arrival models: flows that generate packets periodically with deadline = period, and one-shot arrivals where each flow has a single packet present initially. It is required to multicast these flows to a different multicast group of receivers. Wireless links are error-prone in practice [1]; a result of signal quality degradation due to attenuation with distance, multipath fading and interference. We assume that the wireless link from the AP to each station is an erasure channel, independent across slots and stations. Since wireless link quality is heterogeneous [1], depending on the positions of the transmitter and the receiver, the channel from the AP to receiver $j$ is assumed to have an erasure probability $p_j$. As a special case, we also consider the homogeneous case where $p_j = p \;\; \forall \, j$.

A communication strategy for the AP consists of deciding which packet to broadcast in each slot. We wish to find a strategy that minimizes the expected total deadlines missed across all receivers, where a receiver counts a packet as a miss if it does not receive it by its deadline. Since we are interested in determining the best possible strategy, we assume initially that at the end of every slot, each receiver provides feedback to the transmitter about whether the packet was successfully decoded or not. In other words, *there is complete feedback at the end of every slot*. This problem is a "restless bandit" [2], an extension of the classical multi-armed bandit problem [3] in which bandits not acted upon in a given slot continue to evolve, in contrast to (static) bandits where the state of bandits that are not acted upon in a given slot is frozen. For a relaxation of restless bandits in which the sample path constraint on the number of active bandits per slot is replaced with a constraint on the average number of active bandits per slot, Whittle [2] used a Lagrangian approach to derive a heuristic index policy, provided that the restless bandit problem under consideration satisfies a certain "Whittle indexability" property.

Given that Whittle indexability represents a dividing line between the "easy" and "hard" restless bandit problems, we focus on establishing it for the relaxed multicast real-time scheduling problem. For the Whittle relaxation, we show that for each flow, the AP's decision between transmitting in a slot and idling has a threshold structure. Through the use

Vivek Raghunathan, Min Cao and P. R. Kumar are with the Dept. of Electrical and Computer Eng., University of Illinois, Urbana, USA. Email: {vivek, mincao, prkumar}@csl.uiuc.edu. Vivek Borkar is with the Tata Institute of Fundamental Research, Mumbai, India. Email: borkar@mailhost.tifr.res.in.

This material is based upon work partially supported by NSF under Contract Nos. CCR-0325716 and CNS 05-19535, DARPA/AFOSR under Contract No. F49620-02-1-0325, DARPA under Contact Nos. N00014-0-1-1-0576 and N66001-06-C-2021, Oakridge-Battelle under Contract 239 DOE BATT 4000044522 and AFOSR under Contract No. F49620-02-1-0217. Vivek Raghunathan is also supported by a Vodafone graduate fellowship. Vivek Borkar is supported by a grant from General Motors India Ltd (in addition to others).

of complete feedback, the AP keeps track of the number of receivers $x_i$ that are yet to successfully decode each packet $i$. Then, for the homogeneous case in which the channel to each receiver is identically distributed with erasure probability $p$, the Whittle index corresponding to each flow is $x_i(1-p)$. For the heterogeneous case in which the erasure channel to receiver $j$ has loss probability $p_j$, the Whittle index corresponding to each flow is $\sum_j (1-p_j)$, where the sum ranges over all receivers in the flow's multicast group that are yet to receive the packet.

An interesting question concerns how good the Whittle relaxation performs. To address this, we derive bounds between the performance of the optimal Whittle relaxed scheduler and the optimal exact scheduler. This derivation is based on a novel combination of stochastic dominance and the constraint on the average number of active flows per slot.

In practice, a natural index policy consists of scheduling the flow with the maximum Whittle index in each slot. This policy requires the AP to keep track of the number of receivers $x_i$ that are yet to successfully decode each packet $i$ on a per-slot basis. The use of complete feedback from each receiver in every slot to implement this policy is a strong assumption. This is alleviated by using statistical estimation theory to design a scalable, albeit noisy feedback mechanism that uses only a constant overhead of a single ACK per broadcast packet. The resultant policy is both simple and highly amenable to low-complexity implementation.

In Section II, we describe the problem formulation. In Section III and Section IV, we study the homogeneous and heterogeneous receiver cases. In Section V, we bound the performance of the Whittle relaxation. In Section VI, we relax the complete feedback assumption by designing a noisy feedback low-overhead mechanism.

## II. Model and Formulation as a Restless Bandit

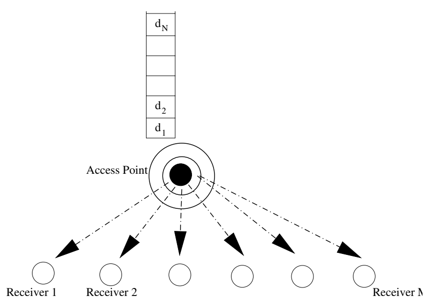


Fig. 1. A wireless real-time broadcast downlink.

Consider a slotted wireless LAN consisting of a wireless AP transmitting real-time information to $M$ collocated receivers, as in Fig. 1. There are $N$ multicast flows. We consider two arrival models: (i) One-shot flows, each comprising one packet each, that arrive at $t=0$ with deadlines $d_i$. (ii) Periodic flows with deadline = period = $t_i$, and initial deadlines $d_i$. Each flow $i$ has a multicast group $G_i$ of receivers interested in its packets. All packets have the same size, equal to one slot.

We model the lossy nature of the wireless medium by associating an erasure channel with erasure probability $p_j$ between the AP and receiver $j$, independent across slots and receivers. The broadcast nature of the medium is modeled as follows: in every slot $k$, the AP selects a packet and broadcasts it on the medium. Each receiver $j$ has an independent probability $1-p_j$ of successfully decoding this packet. This is analogous to the popular ETX metric used to characterize wireless links in practice [1]. A communication strategy for the AP is defined by the sequence of packets broadcast in each slot. We focus on finding a strategy that minimizes the expected total deadlines missed across all receivers, where a receiver counts a packet as missed if it cannot successfully decode it by its deadline.

### A. Formulation as a restless bandit

The problem of determining the optimal strategy is a stochastic dynamic programming problem. We assume complete feedback at the end of every slot, i.e., all receivers provide feedback to the transmitter about whether the packet transmitted was successfully decoded or not. In the homogeneous case where the channel from the AP to each receiver is identically distributed, it suffices to keep track of the number of receivers $x_i$ who are yet to successfully decode the packet $i$. Suppose at some instant $t$, the relative deadlines of the $N$ flows are denoted by the vector $\bar{D} := (D_1, ..., D_N)$. Given the state $((D_1, ..., D_N), (x_1, ..., x_N))$ at time $t$, if the unexpired packet $i$ is scheduled for transmission at time $t$, then each of the $x_i$ receivers who have not yet successfully decoded the packet have an independent Bernoulli probability $(1-p)$ of decoding it. Thus, for $0 \le r \le x_i$, we move into the state $(x_1, ..., r, ..., x_N))$ with a binomial probability $\binom{x_i}{r}(1-p)^r p^{x_i - r}$. Moreover, irrespective of the outcome of the transmission, each of the packets that have not yet expired move one slot closer to their deadline. Thus, for one shot arrivals, the new deadlines are always $((D_1-1)_+, ..., (D_N-1)_+)$, where $(x)_+$ denotes max $(x, 0)$. For periodic arrivals, a new packet from the flow arrives with relative deadline $t_i$ when the previous packet expires, thus the new deadlines are always $((D_1-1)\text{mod } t_1, ..., (D_N-1)\text{mod } t_N)$.

For the general heterogeneous case, we must keep track of the bitmap of receivers that have received each particular packet. To simplify the notation, we focus on the case where there are two groups of receivers. Receivers in group 1 have erasure probability $p_1$, while receivers in group 2 have erasure probability $p_2$. We keep track of the number of receivers $x_{1,i}$ and $x_{2,i}$ in groups 1 and 2 that have received each packet $i$ respectively. Given the state of the system $((D_1, ..., D_N), (x_{1,1}, ..., x_{N,1}), (x_{1,2}, ..., x_{N,2}))$ at time $t$, if the unexpired packet $i$ is scheduled for transmission at time $t$, then each of the $x_{i,k}$ receivers who have not yet successfully decoded packet $i$ have an independent probability $(1-p_k)$ of decoding it, where $k = 1, 2$. Thus, for $0 \le r_k \le x_{i,k}$, we move into the state $(x_{1,1}, ..., r_1, ..., x_{N,1}), (x_{1,2}, ..., r_2, ..., x_{N,2})$ with a probability $\binom{x_{i,1}}{r_1}(1-p_1)^{r_1} p_1^{x_{i,1}-r_1} \times \binom{x_{i,2}}{r_2}(1-p_2)^{r_2} p_2^{x_{i,2}-r_2}$. As in the homogeneous case, each of the packets that have not yet expired move one slot closer to their deadline, and thus, the new deadlines are always $((D_1-1)_+, ..., (D_N-1)_+)$, or $((D_1-1)\text{mod } t_1, ..., (D_N-1)\text{mod } t_N)$, depending on whether the arrival model is one-shot or periodic. The "two receiver group" model generalizes without modification to the general heterogeneous case.

We model deadline misses by using a one-step cost function $c((D_1,...,D_N),(x_1,...,x_N))$. When a packet $i$'s deadline expires, it causes a deadline miss at each of the $x_i$ receivers who are yet to successfully decode it. For the homogeneous special case, $c((D_1,...,D_N),(x_1,...,x_N)) = \sum_{i=1}^{N} x_i \mathbf{1}_{D_i=0}$, where $\mathbf{1}_e$ is thus the indicator function corresponding to the event $e$. Similarly, for the two receiver group heterogeneous case, $c((D_1,...,D_N),(x_{1,1},...,x_{N,1}),(x_{1,2},...,x_{N,2})) = \sum_{i=1}^{N}(x_{1,i}+x_{2,i})\mathbf{1}_{D_i=0}$.

We use restless bandits [2] from stochastic control to study this problem. In the contrasting classical static multi-armed bandit problem [3], bandits that are not acted upon in a slot are "frozen". The restless bandits problem [2] is an extension where the bandits that are not acted upon in a slot continue to evolve. Optimal multicast real-time scheduling is a restless bandit. For homogeneous receivers, represent flow $i$'s state by $(D_i, x_i)$. For $D_i > 0$, if flow $i$ is scheduled in a particular slot, it moves into the state $(D_i - 1, r)$ with probability $\binom{x_i}{r} p^r (1-p)^{x_i - r}$. When $D_i = 0$, with periodic arrivals, it moves into the state $(t_i, |G_i|)$, where $|G_i|$ is the size of the multicast group $G_i$. For one shot arrivals, it moves into the state $(0,0)$. In both cases, a one-shot cost equal to the number of outstanding receivers is incurred at $D_i = 0$. In the language of [2], this is the "active" evolution of the flow. On the other hand, if flow $i$ is not scheduled in a particular slot and $D_i > 0$, it moves into the state $(D_i - 1, x_i)$. If $D_i = 0$, the evolution of the state is identical to the "active" case. This is the "passive" evolution of the flow.

The general restless bandits problem is PSPACE-COMPLETE [4]. Whittle [2] proposed a relaxation to the restless bandits problem in which the *sample path constraint* on the number of active flows (bandits) per slot is replaced by *a constraint on the expected number* of active flows per slot. With this relaxation, a Lagrangian formulation decouples the problem into $N$ one-dimensional problems, one for each flow, each consisting of determining whether or not a flow should be active in a slot given a reward of $-\alpha$ per slot in which it is passive. These sub-problems are only coupled through the Lagrange multiplier $-\alpha$, and the optimal $-\alpha^*$ corresponds to the "passivity reward" that ensures that the constraint on the average number of active flows being one is met.

A restless bandit problem is Whittle indexable [2] if for each bandit, the set of states for which it is optimal to stay passive increases monotonically as the reward $-\alpha$ increases. Whittle indexability allows one to consistently define and compute an index for each bandit. A natural index policy consists of scheduling the flow with the maximum Whittle index in each slot. Since such an index policy is derivable only under the assumption of Whittle indexability, this property represents somewhat of a dividing line between the "easy" and "hard" restless bandit problems. We focus on establishing indexability and computing the Whittle index for the relaxation of real-time multicast scheduling.

## III. Building Intuition: Homogeneous Channels

We start off by showing some monotonicity properties for both one shot and periodic arrivals.

*Lemma 1:* Consider two potential starting states, the first with a state where the vector of the number of outstanding receivers for the flows is $(a_1, a_2, ..., a_N)$, and the second where the vector of the number is $(b_1, b_2, ..., b_N)$, with $b_i \leq a_i$. For one shot arrivals, the optimal policy for the exact wireless broadcast real-time scheduling problem starting from the second state has no greater deadline misses than the optimal policy starting from the first state.

*Proof:* (Outline) The idea is to use stochastic dominance combined with a double induction argument, the outer induction on the number of flows, and the inner induction backward in time. Observe that a binomial random variable $Y$ with parameter $(x_1+1, p)$ stochastically dominates a binomial random variable $X$ with parameter $(x_1, p)$. Also observe that if $Y$ dominates $X$ and $f$ is monotonic increasing, then $E_Y[f] \geq E_X[f]$. We use these two facts to establish monotonicity for a single flow by inducting backward in time. Then, assuming that monotonicity holds when there are $k-1$ flows in the system, we repeat stochastic dominance to complete the proof. See [5] for details. ■

*Lemma 2:* Consider two potential starting states, the first with a state where the vector of the number of outstanding receivers for the flows is $\{a_1, a_2, ..., a_N\}$, and the second where the vector of the number is $\{b_1, b_2, ..., b_N\}$, with $b_i \leq a_i$. For periodic arrivals, the optimal policy for the exact wireless broadcast real-time scheduling problem starting from the second state has no greater deadline misses than the optimal policy starting from the first state.

*Proof:* (Outline) The proof is similar to Lemma 1, and uses the same stochastic dominance argument. The only difference is the use of value iteration, inducting on the number of iterations $n$ of the dynamic programming operator $T$ [6], instead of backward induction. See [5] for details. ■

We can use the monotonicity to show that the optimal policy for the exact scheduling problem is non-idling.

*Lemma 3:* For both one shot and periodic arrivals, there exists an optimal policy of the real time broadcast scheduling problem that is non-idling.

*Proof:* The proof is by stochastic dominance. Consider an optimal policy that idles in a set of states $S$. Consider the stationary non-idling policy obtained by emulating the optimal policy when not in $S$, and scheduling the earliest deadline first packet when in $S$. Suppose the sample paths are coupled so that the binomial random variables have the same realization for both the optimal policy and the modified policy. On every sample path corresponding to the optimal policy, consider the first instant $t$ at which the policy is in the set of states $S$. Since the policy idles at time $t$, the number of outstanding receivers at time $t+1$ is the same as that at time $t$. On the other hand, the modified policy schedules the same flows as the optimal policy till time $t$, and schedules the earliest deadline packet at time $t+1$. Then, at $t+1$, the number of outstanding receivers for the modified policy is lesser than or equal to the optimal policy for each flow on each sample path. By monotonicity (Lemma 1, Lemma 2), the cost-to-go for the modified policy from $t+1$ is lesser than or equal to the cost-to-go for the optimal policy, and we have constructed a non-idling optimal policy. ■

As described in Section II, for the multicast real-time scheduling problem under consideration, in every slot, flows continue to move one slot closer to their deadline whether or not the flow is scheduled in that slot. This makes multicast real-time scheduling a restless bandits problem. We now consider the Whittle relaxation of the real time scheduling problem where the *sample path constraint* of one active flow per slot is replaced with a relaxed constraint on the average number of flows per slot being one. Under this relaxation, we can use a Lagrangian approach to decouple the multi-dimensional scheduling problem into $N$ one-dimensional stochastic control problems for each of the flows (bandits). These sub-problems interact with each other only through a scalar Lagrange multiplier $\alpha \leq 0$. Each of these sub-problems corresponds to determining whether or not the flow should stay active in a slot, where a flow is given a reward of $-\alpha$ per slot for staying passive in any slot. In other words, the scalar Lagrange multiplier $\alpha$ can be interpreted as a cost for passivity, and the optimal $-\alpha^*$ is the value of passivity reward that ensures that the relaxed constraint of the average number of active flows being one is met.

Define $V^\alpha(x_i, d)$ as the optimal expected cost-to-go for a single flow where $\alpha \leq 0$ is the cost of passivity, $x_i$ is the number of outstanding receivers, and $d$ is the time to deadline. Also, define $P_\alpha$ as the set of states $(x, d)$ for which it is optimal for a flow to idle (stay passive) instead of transmitting when the reward is $-\alpha$. Then, a restless bandit problem is said to be Whittle indexable [2] if for each bandit, $P_\alpha$ increases monotonically as the reward $-\alpha$ increases. For restless bandits that satisfy the Whittle indexability property, it is possible to consistently define and associate an index with each bandit. The computation of this index allows for the derivation of a natural index policy that schedules the bandit with the maximum index in each slot.

We now establish an important lower bound that will aid in establishing Whittle indexability for the homogeneous case.

*Lemma 4:* For a single flow in the Whittle relaxation, the optimal cost-to-go function $V^\alpha(x, d)$ satisfies:

$$V^\alpha(x,d) - \sum_{y=0}^{x} \binom{x}{y} p^y (1-p)^{x-y} V^\alpha(y,d) \geq (1-p_h)(-\alpha) + p_h x(1-p), \qquad (1)$$

where $p_h$ is a deterministic binary $0-1$ variable that is 1 if the optimal time varying state feedback policy $u(x,t)$ is passive for all $t \in [0,d]$ starting from $(x,d)$ down to $(x,0)$, and 0 otherwise. This is true for both periodic and one shot arrivals.

*Proof:* The first term on the LHS is the cost of playing the optimal policy for the first $d$ slots. Observe that the second term on the LHS corresponds to the expected cost-to-go starting from state $(x, d+1)$ and playing active for a single slot. Note that the term $V^\alpha(y,d)$ represents the cost of the optimal policy from state $y$ with $d$ slots till deadline. Hence, if we replace it with a non-optimal policy from state $y$ onwards, then the difference between the first and second terms could only have decreased. If we can show that this diminished difference still exceeds the RHS, then the result is proved. This is what we shall do.

Consider first the case where $p_h = 1$, i.e., the first term on the LHS corresponds to staying inactive for all $d$ slots, and collecting a reward of $-\alpha \cdot d$, and paying $x$ for the number of jobs with expired deadlines. Now, let us replace the $V^\alpha(y,d)$ term on the LHS by a suboptimal and larger cost of just remaining passive for all $d$ slots. This policy too collects a reward of $-\alpha.d$ for remaining passive for $d$ slots. However, the expected value of the number of jobs left after the first active slot is only $x.p$. Hence, it only pays a penalty of $x.p$ for the expected number of deadline misses. Thus, the difference between the first and second terms on the LHS is no less than $x - xp = x(1-p)$. This is the term $p_h x(1-p)$ on the RHS.

Now, suppose that the optimal policy for the first term on the LHS does not remain passive for all $d$ slots, but instead it remains passive only for $t$ consecutive slots, and then becomes active. In that case, the value of $t$ is known a priori. Now, in the second term, again consider a suboptimal policy that from the state $y$ (cost-to-go represented by $V^\alpha(y,d)$) remains passive for $t+1$ slots, and then gets active with its evolution stochastically coupled to the same path for the first term after its $t$ passive slots, as shown in Fig. 2. Both will end with the same number of deadline misses, but the second term has been replaced by something that is active for one more slot, namely slot $t$, than the first term, thus collecting an extra reward of $-\alpha$. This is the term $(1-p_h)(-\alpha)$ on the RHS.

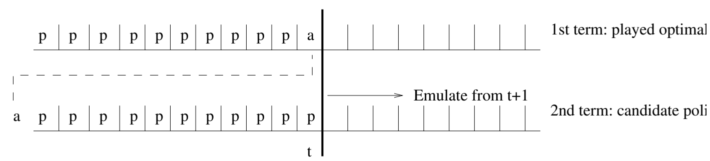


Fig. 2. Proof of Lemma 4.

■

In fact, Lemma 4 can be generalized to other evolutions of the flow in the active state, as captured by the following lemma, stated without proof (see [5] for details).

*Lemma 5:* Consider the Whittle relaxation with passivity reward $-\alpha$, for a modified scheduling problem in which, when it is active in a slot, a flow in state $(x,d)$ evolves according to transition probability $S(x,y)$. Then, for a single flow in this Whittle relaxation with this modified dynamics, the optimal cost-to-go function $V^\alpha(x,d)$ satisfies:

$$V^\alpha(x,d) - \sum_{y=0}^{n} S(x,y) V^\alpha(y,d) \geq (1-p_h)(-\alpha) + p_h(x - \sum_{y=0}^{n} y.S(x,y)), \qquad (2)$$

where $p_h$ is a deterministic binary $0-1$ variable that is one if the optimal policy $u(x,t)$ for this modified problem is passive starting from $(x,d)$ down to $(x,0)$, and zero otherwise. This is true for both periodic and one shot arrivals.

We now use a different policy emulation argument to establish an important upper bound that will also aid in establishing Whittle indexability for the homogeneous case.

*Lemma 6:* For a single flow in the Whittle relaxation, the cost-to-go function $V^\alpha(x,d)$ satisfies:

$$V^\alpha(x,d) - \sum_{y=0}^{x} \binom{x}{y} p^y (1-p)^{x-y} V^\alpha(y,d) \leq (1-p_l)(-\alpha) + p_l x(1-p), \quad (3)$$

where $p_l$ is the probability that the optimal policy is active on all slots from $t = d$ down to $t = 0$. This is true for both periodic and one shot arrivals.

*Proof:* Observe that the second term on the LHS corresponds to starting from state $(x, d+1)$, playing active for a single slot, followed by playing optimally for the remaining $d$ slots. Call the resulting sequence of actions $A$. Instead of the first term in the LHS, consider the following candidate policy starting from $(x,d)$ that not being necessarily optimal leads to a larger cost: If $A$ has an initial segment of $t$ consecutive active plays followed by a passive play, then schedule $t+1$ consecutive active plays and thereafter copy what $A$ does. This policy can be implemented assuming that the sample paths are stochastically coupled. Call the resulting action sequence $B$. Since we consider a candidate policy for the $1^{st}$ term and the optimal policy for the $2^{nd}$ term, we obtain an upper bound.

If $t$ is not $d$, that is, if B does have an extra passive play compared to A, shown in Fig. 3, then B collects an extra reward of $-\alpha$. This is the term $(1-p_l)(-\alpha)$ on the RHS.

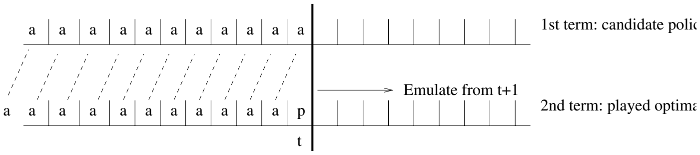


Fig. 3. Proof of Lemma 6.

The only sample paths on which such an emulation is not possible are sample paths in which $t = d$, i.e., the optimal policy for the second term on the LHS is active on all slots till expiry. Suppose the probability of this event is $p_l$. In this case, the LHS is upper bounded by the difference in the cost incurred by staying active from $(x,d)$ continuously till expiry versus the cost incurred by staying active from $(x, d+1)$ continuously till expiry. Observe that this difference is only calculated over sample paths on which it was optimal to be active on all slots till expiry. This implicit conditioning makes the calculation of this difference quite subtle. Consider any sample path for which the optimal policy for the second term on the LHS is active on all slots till expiry. Suppose the number of outstanding receivers in slot $t$ is $x_t$. Then, the expected number of deadline misses for the second term on the LHS is $x_d p$. The first term in the LHS is stochastically coupled to the second term and is always "one slot behind" the second term in its evolution, as shown in Fig. 3. Since we coupled the realizations of the two terms, the number of deadline misses for the first term is in fact $x_d$. Thus, the difference in the two terms is given by $x_d - x_d p = x_d(1-p) \leq x(1-p)$. This is the term $p_l x(1-p)$ on the RHS. ■

As with Lemma 4, the proof of Lemma 6 can be generalized to other one-slot evolutions of the flow (bandit) in the active state, captured in the following lemma, stated without proof (for details, see [5]).

*Lemma 7:* Consider the Whittle relaxation with passivity reward $-\alpha$, for a modified scheduling problem in which, when it is active in a slot, a flow in state $(x,d)$ evolves according to transition probability $S(x,y)$. Then, for a single flow in this Whittle relaxation with this modified dynamics, the optimal cost-to-go function $V^\alpha(x,d)$ satisfies:

$$V^\alpha(x,d) - \sum_{y=0}^{n} S(x,y) V^\alpha(y,d) \leq (1-p_l)(-\alpha) + p_l(E_d - E_{d+1}). \quad (4)$$

Above, $p_l$ is the probability that the optimal policy for this modified problem is active on all slots from $t = d-1$ down to $t = 0$ starting from $(x,d)$, and $E_t$ is the expected number of outstanding receivers under $S(x,y)$ when the flow is scheduled for $t$ consecutive slots starting from $(x,d)$, where the expectation is conditioned over sample paths for which it is optimal to stay active on all slots till expiry. This is true for both periodic and one shot arrivals.

We are now ready to establish a threshold structure for the optimal policy for a single bandit in the Whittle relaxation.

*Theorem 1:* (Threshold structure) Consider a single bandit in the Whittle relaxation of the wireless real-time broadcast scheduling problem. Suppose $\alpha \leq 0$ is the Lagrange multiplier representing the cost for passivity. Then, there exists an optimal policy $u^\alpha(x,d)$ with the following threshold structure (both decisions being optimal when there is equality):

$$x \geq \frac{-\alpha}{1-p} \Rightarrow u^\alpha(x,d) = \text{active},$$
$$x \leq \frac{-\alpha}{1-p} \Rightarrow u^\alpha(x,d) = \text{passive}.$$

*Proof:* Let $x \geq \frac{-\alpha}{1-p}$. Then, from Lemma 4, we have:

$$V^\alpha(x,d-1) - \sum_{y=0}^{x} \binom{x}{y} p^y (1-p)^{x-y} V^\alpha(y,d-1) \geq -\alpha + p_h(\alpha + x(1-p)) \geq -\alpha. \quad (5)$$

Let $x \leq \frac{-\alpha}{1-p}$. Then, from Lemma 6, we have:

$$V^\alpha(x,d-1) - \sum_{y=0}^{x} \binom{x}{y} p^y (1-p)^{x-y} V^\alpha(y,d-1) \leq -\alpha + p_l(\alpha + x(1-p)) \leq -\alpha. \quad (6)$$

Combining (5) and (6) establishes the existence of an optimal $u(x,d)$ which uses an active action whenever $x \geq \frac{-\alpha}{1-p}$, and a passive action whenever $x \leq \frac{-\alpha}{1-p}$, shown in Fig. 4. ■

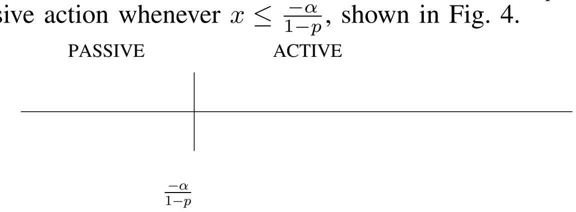


Fig. 4. Threshold structure for optimal policy.

We are now ready to prove the Whittle indexability of the wireless real-time broadcast scheduling problem.

*Theorem 2:* Consider a single bandit in the Whittle relaxation of the wireless real-time broadcast scheduling problem. Suppose $\alpha \leq 0$ is the Lagrange multiplier representing the cost for passivity. Define $P_\alpha$ to be the set of all $(x,d)$ such that it is optimal to idle (stay passive) instead of transmitting (active). Then,

$$(x, d) \in P_\alpha \Rightarrow (x, d) \in P_{\alpha'} \quad \forall \quad \alpha' \leq \alpha. \tag{7}$$

In other words, the wireless broadcast real-time scheduling problem is Whittle-indexable, since increasing the reward for passivity increases the tendency to stay passive.

*Proof:* Given $\alpha \leq 0$, we know that the policy $u(x, d)$ which uses an active action whenever $x \geq \frac{-\alpha}{1-p}$ and passive action otherwise, is an optimal policy. Suppose $(x, d) \in P_\alpha$. Then, $x \leq \frac{-\alpha}{1-p} \leq \frac{-\alpha'}{1-p} \quad \forall \ \alpha' \leq \alpha$. This implies that $(x, d) \in P_{\alpha'}$. This establishes Whittle indexability, as the passive set $P_\alpha$ grows monotonically as $-\alpha$ goes from $-\infty$ to $\infty$. ∎

We have thus established that the Whittle relaxation of wireless multicast real-time scheduling is indeed Whittle indexable. In [2], the Whittle index of a state is defined as that value of the passivity reward (Lagrange multiplier) that makes a bandit indifferent to being passive or active. Thus, the Whittle index for a flow in state $(x, d)$ is $x(1-p)$. We can use this Whittle index to derive an index policy for wireless real-time scheduling of multicast applications.

**Whittle index policy for multicast real-time scheduling**: In every slot, schedule the unexpired flow with the maximum $x(1-p)$.

## IV. Heterogeneous Channels

We now establish Whittle indexability for the case with heterogeneous receivers, and derive the corresponding Whittle index policy. For simplicity of notation, we focus on the two receiver groups case described in Section II. We must emphasize that our results generalize without modification to the general heterogeneous case. In the two receiver groups case, the state of any flow is captured by the three tuple $(x_1, x_2, d)$, where $x_1$ is the outstanding receivers of group 1, $x_2$ is the outstanding receivers of group 2, and $d$ is the slots to expiry. As before, the Whittle relaxation decouples the global problem of deciding which of $N$ flows to schedule, into $N$ simpler sub-problems of deciding whether a flow should stay active or passive given a reward of $-\alpha$ per time slot for passivity, where $-\alpha$ is the Lagrange multiplier coupling the sub-problems. We start off by establishing lower and upper bounds similar to Lemma 4 and Lemma 6.

*Lemma 8:* For a single flow in the Whittle relaxation, the cost-to-go function $V^\alpha(x_1, x_2, d)$ satisfies:

$$V^\alpha(x_1, x_2, d) - \sum_{y_1=0}^{x_1} \sum_{y_2=0}^{x_2} \binom{x_1}{y_1} p_1^{y_1}(1-p_1)^{x_1-y_1} \times \binom{x_2}{y_2} p_2^{y_2}(1-p_2)^{x_2-y_2} V^\alpha(y_1, y_2, d) \geq (1-p_h)(-\alpha) + p_h[x_1(1-p_1) + x_2(1-p_2)],$$

where $p_h$ is a deterministic binary $0-1$ variable that is $1$ if the optimal time varying state feedback policy is passive for all $t \in [0, d]$ starting from $(x_1, x_2, d)$ down to $(x_1, x_2, 0)$, and $0$ otherwise. This is true for both periodic and one shot arrivals.

*Proof:* (Outline) The proof is similar to the proof of Lemma 4; the difference is in the calculations. For the first term on the LHS, consider playing the optimal policy for the first $d$ slots. Consider the candidate policy for the second term that stays passive in every slot till the first time $t$ (known apriori) such that the optimal policy from the state $(x_1, x_2, t)$ is active. The candidate policy stays passive at time $t$ and then emulates the optimal policy starting from $t+1$ onwards.

Consider a coupled realization of the two systems where for each receiver group, the realization of the binomial random variables is the same for both terms, time shifted forward for the first term. First, consider states where $p_h = 0$, in which case the candidate policy is able to stay passive for an extra time slot. Stochastic coupling then ensures that the difference in the first and second terms in the LHS is lower bounded by the reward $-\alpha$ for staying passive for an extra slot.

The only sample paths on which such an emulation is not possible are those in which the optimal policy for the first term on the LHS is to be passive on all slots till expiry. On such sample paths, $p_h$ is 1, and the LHS is lower bounded by the difference in the cost incurred by staying passive continuously for both terms till expiry, i.e., $x_1(1-p_1) + x_2(1-p_2)$. ∎

*Lemma 9:* For a single flow in the Whittle relaxation, the cost-to-go function $V^\alpha(x_1, x_2, d)$ satisfies:

$$V^\alpha(x_1, x_2, d) - \sum_{y_1=0}^{x_1} \sum_{y_2=0}^{x_2} \binom{x_1}{y_1} p_1^{y_1}(1-p_1)^{x_1-y_1} \times \binom{x_2}{y_2} p_2^{y_2}(1-p_2)^{x_2-y_2} V^\alpha(y_1, y_2, d) \leq (1-p_l)(-\alpha) + p_l[x_1(1-p_1) + x_2(1-p_2)],$$

where $p_l$ is the probability that the optimal policy is active on all slots from $t = d$ down to $t = 0$. This is true for both periodic and one shot arrivals.

*Proof:* (Outline) The proof is similar to the proof of Lemma 6; the only difference is in the calculations. For the second term on the LHS, play the optimal policy for the first $d$ slots; the cost-to-go for the second term after $t$ slots thus corresponds to being active for a single slot from $(x_1, x_2, d+1)$ and then being optimal for the next $t$ slots. For the first term, consider the following candidate policy starting from $(x_1, x_2, d)$: stay active in every slot till the first time $t$ such that the optimal policy from the state at time $t$ is passive. Continue to stay active at time $t$ and then emulate the optimal policy starting from time $t+1$ onwards till expiry.

Consider a coupled realization of the two systems where for each receiver group, the binomial random variables have the same realization for both terms, time shifted forward by one unit for the first term till time $t$. On sample paths where the candidate policy is able to stay active for an extra slot in such a manner, a stochastic coupling argument ensures that the difference in the cost-to-go for the first and second terms is upper bounded by the reward for staying passive for an extra unit for the second term = $-\alpha$.

The sample paths on which such an emulation is not possible are those in which the optimal policy for the second term on the LHS is active on all slots till expiry. Here, the LHS is upper bounded by the difference in the cost incurred by staying active from $(x_1, x_2, d)$ continuously till expiry versus the cost incurred by staying active from $(x_1, x_2, d+1)$ continuously till expiry. This difference is only calculated over sample paths on which it is optimal to be active on all slots till expiry. Suppose the number of outstanding receivers in slot $t$ is $(x_{t,1}, x_{t,2})$. Then, the expected number of deadline misses for the second term on the LHS is $x_{d,1}p_1 + x_{d,1}p_2$. Since the realizations were stochastically coupled and the first term is always "one slot behind" the second term in its evolution, the number of deadline misses for the first term is in fact $x_{d,1} + x_{d,2}$. This difference is $x_{d,1}(1-p_1)+x_{d,2}(1-p_2) \leq x_1(1-p_1)+x_2(1-p_2)$. ■

We now establish threshold structure for the optimal Whittle relaxed policy for each flow.

*Theorem 3:* (Threshold structure) Consider a single "bandit" in the Whittle relaxation of the wireless real-time broadcast scheduling problem with two groups of receivers. Suppose $\alpha \leq 0$ is the Lagrange multiplier representing the cost for passivity. Then, there exists an optimal policy $u^\alpha(x_1, x_2, d)$ with the following threshold structure (both decisions being optimal when there is equality):

$$x_1(1-p_1) + x_2(1-p_2) > -\alpha \Rightarrow u^\alpha(x_1, x_2, d) = \text{active},$$
$$x_1(1-p_1) + x_2(1-p_2) < -\alpha \Rightarrow u^\alpha(x_1, x_2, d) = \text{passive}.$$

*Proof:* Let $x_1(1-p_1)+x_2(1-p_2) \geq -\alpha$. From Lemma 8, we have:

$$V^\alpha(x_1, x_2, d-1) - \sum_{y_1=0}^{x_1}\sum_{y_2=0}^{x_2} \binom{x_1}{y_1} p_1^{y_1}(1-p_1)^{x_1-y_1} \times \binom{x_2}{y_2} p_2^{y_2}(1-p_2)^{x_2-y_2} V^\alpha(y_1, y_2, d-1) \geq -\alpha + p_h(\alpha + [x_1(1-p_1) + x_2(1-p_2)]) \geq -\alpha. \quad (8)$$

Let $x_1(1-p_1) + x_2(1-p_2) \leq -\alpha$. From Lemma 9, we have:

$$V^\alpha(x_1, x_2, d-1) - \sum_{y_1=0}^{x_1}\sum_{y_2=0}^{x_2} \binom{x_1}{y_1} p_1^{y_1}(1-p_1)^{x_1-y_1} \times \binom{x_2}{y_2} p_2^{y_2}(1-p_2)^{x_2-y_2} V^\alpha(y_1, y_2, d-1) \leq -\alpha + p_l(\alpha + [x_1(1-p_1) + x_2(1-p_2)]) \leq -\alpha. \quad (9)$$

Combining (8) and (9) establishes the existence of an optimal $u(x_1, x_2, d)$ with the structure described in Thm. 3, as shown in Fig. 5. ■

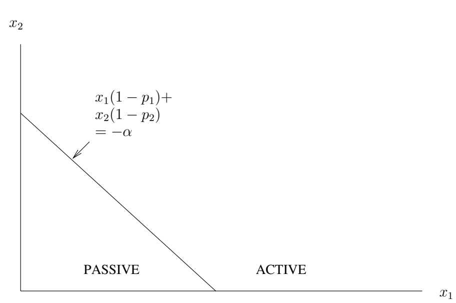


Fig. 5. Threshold structure for optimal policy with two receiver groups.

We are now ready to prove the Whittle indexability of the wireless real-time broadcast scheduling problem with two receiver groups.

*Theorem 4:* Consider a single bandit in the Whittle relaxation of the wireless real-time broadcast scheduling problem with two receiver groups. Suppose $\alpha \leq 0$ is the Lagrange multiplier representing the cost for passivity. Define $P_\alpha$ to be the set of all $(x_1, x_2, d)$ such that it is optimal to idle (stay passive) instead of transmitting (active). Then,

$$(x_1, x_2, d) \in P_\alpha \Rightarrow (x_1, x_2, d) \in P_{\alpha'} \quad \forall \quad \alpha' \leq \alpha. \quad (10)$$

In other words, the wireless broadcast real-time scheduling problem is Whittle-indexable, since increasing the reward for passivity increases the tendency to stay passive.

*Proof:* Given $\alpha \leq 0$, we know that the policy $u(x_1, x_2, d)$ which uses an active action whenever $x_1(1-p_1)+x_2(1-p_2 \geq -\alpha$ and passive action otherwise, is an optimal policy. Suppose $(x_1, x_2, d) \in P_\alpha$. Then, $x_1(1-p_1) + x_2(1-p_2) \leq -\alpha \leq -\alpha' \quad \forall \ \alpha' \leq \alpha$. This implies that $(x_1, x_2, d) \in P_{\alpha'}$. This establishes Whittle indexability, as the passive set $P_\alpha$ grows monotonically as $-\alpha$ goes from $-\infty$ to $\infty$. ■

For the two receiver group case, we have thus established Whittle indexability for the relaxation. In [2], the Whittle index of a state is defined as that value of the passivity reward (Lagrange multiplier) that makes a bandit indifferent to being passive or active. Thus, the Whittle index for a flow in state $(x_1, x_2, d)$ is $x_1(1-p_1)+x_2(1-p_2)$. We can use this to derive an index policy:

**Whittle index policy for multicast real-time scheduling with two receiver groups**: In every slot, schedule the unexpired flow with the maximum $x_1(1-p_1) + x_2(1-p_2)$.

These arguments extend directly to the general heterogeneous case, where the Whittle index can be computed as $\sum_j (1-p_j)$, where the sum is over all outstanding receivers, and provides an index policy:

**Whittle index policy for multicast real-time scheduling with heterogeneous receivers**: In every slot, schedule the unexpired flow with the maximum $\sum_{j \text{outstanding}} (1-p_j)$.

## V. How well does the Whittle relaxation perform?

We now compute performance bounds for the optimal Whittle relaxed policy with respect to the optimal policy for the exact problem for the special case where all multicast groups have the same size, i.e., $|G_i| = M$. (The calculations for the general case are similar.) Let $\alpha^*$ be the Lagrange multiplier corresponding to the optimum Whittle relaxed policy. Then, for each flow, the optimal policy in the Whittle relaxation consists of scheduling the flow $i$ till $x_i(1-p)$ drops below $-\alpha^*$, or deadline expiry. To bound performance, we will need to bound $\alpha^*$.

Let the random variable $h_i$ denote the time taken for the number of outstanding receivers to decrease from $|G_i|$ to $\frac{-\alpha^*}{1-p}$. The optimal scheduler for the Whittle relaxation schedules the flow $i$ in every such slot, and thus the expected fraction of time for which flow $i$ is active is given by $\frac{E[h_i]}{t_i}$. Summed over all flows, the expected number of active flows per slot is at most one; in fact exactly one by a stochastic dominance argument. Thus, we have:

$$\sum_{i=1}^{N} \frac{E[h_i]}{t_i} = 1. \quad (11)$$

Consider the Markov chain defined on $\{0, 1, ..., |G_i|\}$ with transition probability from a state $x$ to any state $y \leq x$ given by the binomial distribution probability $\binom{x}{y}p^y(1-p)^{x-y}$. Observe that $h_i$ is the hitting time for this Markov chain to the state $\frac{-\alpha^*}{1-p}$ starting from the state $|G_i|$. Denoting by $\psi_j$ the successive jumps of this Markov chain starting from state $|G_i|$, we have, for each sample path $w$:

$$\sum_{j=1}^{h_i(w)} \psi_j(w) = |G_i| - \frac{-\alpha^*}{1-p} \tag{12}$$

### A. Lower bound on $-\alpha^*$

Since $h_i$ is a stopping time, we can use a martingale stopping argument to derive:

$$E[\sum_{j=1}^{h_i} \psi_j] = E[\sum_{j=1}^{h_i}[|G_i|(1-p)p^{i-1}] \geq |G_i| - \frac{-\alpha^*}{1-p}$$

$$\text{Hence, } E[|G_i|(1-p)\frac{1-p^{h_i}}{(1-p)}] \geq |G_i| - \frac{-\alpha^*}{(1-p)}$$

$$\text{and thus, } E[p^{h_i}] \leq \frac{-\alpha^*}{(1-p)|G_i|}.$$

Since $p^x$ is convex, by Jensen's inequality, we have $p^{E[h_i]} \leq E[p^{h_i}]$ and thus,

$$E[h_i] \geq \frac{\log(\frac{-\alpha^*}{(1-p)|G_i|})}{\log p} \tag{13}$$

Combining (11) and (13), we get a lower bound on the optimal Lagrange multiplier $-\alpha^*$:

$$\sum_{i=1}^{N} \frac{\log(\frac{-\alpha^*}{(1-p)|G_i|})}{t_i \log p} \leq 1$$

$$\Rightarrow -\alpha^* \geq p^{\frac{1}{\sum_{i=1}^N \frac{1}{t_i}}} \cdot (1-p) \cdot \prod_{i=1}^{N} |G_i|^{\frac{\frac{1}{t_i}}{\sum_{i=1}^N \frac{1}{t_i}}}$$

When all $|G_i| = M$, this simplifies to:

$$-\alpha^* \geq m_l \triangleq p^{\frac{1}{\sum_{i=1}^N \frac{1}{t_i}}} \cdot (1-p)M \tag{14}$$

### B. Upper bound on $-\alpha^*$

Next, we replace each of the random variables $\psi_j(w)$ with independent draws $\psi_j^l(w)$ from the binomial distribution corresponding to $\frac{-\alpha^*}{1-p}$ number of Bernoulli trials, each with success probability $1-p$. Note that since $G_i$ is no smaller than $\frac{-\alpha^*}{1-p}$, each of the $\psi_j$ stochastically dominates the $\psi_j^l$, and the expected hitting time $E[h_i^l]$ obtained by using the stochastically smaller random variables is an upper bound on $E[h_i]$. Since the $\psi_j^l$ are independent and identically distributed, we can use Wald's equation and (12) to compute the expected hitting time $E[h_i^l]$:

$$E[h_i^l] \times \frac{-\alpha^*}{1-p} \cdot (1-p) \leq |G_i| - \frac{-\alpha^*}{1-p}$$

$$\Rightarrow E[h_i] \leq E[h_i^l] \leq \frac{|G_i| - \frac{-\alpha^*}{1-p}}{-\alpha^*} \tag{15}$$

Combining (11) and (15), we get an upper bound on the optimal Lagrange multiplier $-\alpha^*$:

$$\sum_{i=1}^{N} \frac{|G_i| - \frac{-\alpha^*}{1-p}}{(-\alpha^*)t_i} \geq 1 \Rightarrow -\alpha^* \leq \frac{\sum_{i=1}^N \frac{|G_i|}{t_i}}{1 + \frac{1}{1-p}\sum_{i=1}^N \frac{1}{t_i}}$$

When all $|G_i| = M$, this simplifies to:

$$-\alpha^* \leq m_u \triangleq \frac{M}{\frac{1}{\sum_{i=1}^N \frac{1}{t_i}} + \frac{1}{1-p}} \tag{16}$$

### C. Cost for the optimal policy for the Whittle relaxation

For each flow $i$, the cost of the optimal policy for the Whittle relaxation $V_{i,r}^*$ is lower bounded by $\frac{-\alpha^*}{1-p} + \alpha^*(t_i - E[h_i])_+$. The first term comes from the fact that we stop scheduling the flow once the number of outstanding receivers is $\frac{-\alpha^*}{1-p}$ and thus incur deadline misses for (at least) each of these outstanding receivers. The second term comes from the fact that the flow $i$ is idle once it reaches the state $\frac{-\alpha^*}{1-p}$, i.e., for $t_i - E[h_i]$ slots and for every such slot in which it is idle, it collects a "passivity reward" $-\alpha^*$. From (13), (14), (16), we have:

$$\begin{aligned} V_r^* &= \sum_{i=1}^{N} V_{i,r}^* \geq \sum_{i=1}^{N} \frac{-\alpha^*}{1-p} + \alpha^*(t_i - E[h_i]) \\ &\geq \sum_{i=1}^{N} \frac{-\alpha^*}{1-p} + \alpha^*(t_i - \frac{\log(\frac{-\alpha^*}{M(1-p)})}{\log p}), \text{ since } \alpha^* \leq 0 \\ &\geq \frac{Nm_l}{1-p} - m_u \sum_{i=1}^{N} t_i + \frac{Nm_u}{\sum_{i=1}^N \frac{1}{t_i}}, \text{ since } E[h_i] \leq t_i \end{aligned} \tag{17}$$

### D. Cost for a candidate policy for the exact problem

Consider a candidate policy $P$ that gives each flow $i$ exactly $z_i = \frac{1}{\sum_{i=1}^N \frac{1}{t_i}}$ slots per period $t_i$ of the flow. Assuming that the flows are feasible in the hard real-time sense, i.e., $\sum_{i=1}^N \frac{1}{t_i} \leq 1$, this can be achieved by emulating the schedule generated by an optimal hard real-time scheduler like EDF (ignoring losses) [7]. Then, the expected total deadline misses $V_e^c$ is given by:

$$V_e^c = \sum_{i=1}^{N} Mp^{z_i} = NMp^{\frac{1}{\sum_{i=1}^N \frac{1}{t_i}}},$$

From (14) and the fact that $P$ is a feasible policy for the exact problem, we get:

$$V_e^* \leq V_e^c \leq \frac{Nm_l}{(1-p)} \tag{18}$$

### E. Performance bound on the Whittle relaxation

Combining (17) and (18), we get a closed form bound:

$$V_e^* - V_r^* \leq m_u \sum_{i=1}^{N} t_i - \frac{Nm_u}{\sum_{i=1}^N \frac{1}{t_i}}$$

## VI. A Statistical Estimation Technique for Scalable Receiver Feedback

The results of Section III and Section IV were derived under the assumption of complete feedback. At the end of every packet transmission, we required all outstanding receivers to provide the AP information on whether or not they successfully received the packet. This can be implemented by having each receiver communicate an ACK at the end of every transmission to the AP, indicating whether or not it received the transmitted packet. This is overhead-prone, specially when the number of receivers is large. This is exacerbated for 802.11-like wireless systems in which: (i) PHY overhead is considerable, and (ii) channel access is mediated using random access. (In practice, IEEE 802.11 disables ACKs for broadcast/multicast transmission.)

We now outline a simple mechanism to provide feedback for multicast transmissions on a broadcast wireless downlink. This mechanism is completely distributed, with a bounded communication overhead per broadcast transmission, irrespective of the number of receivers. It incurs a certain amount of delay per transmission; this delay is controllable and can be traded off for estimation error. We envisage the estimates from the noisy feedback mechanism being used in the index policy as if they were the true values.

The mechanism works as follows: if a receiver $j$ decodes a packet $i$ from the AP, and $j$ is an outstanding receiver in the multicast group $G_i$, it sets a random timer $X_j$ distributed according to the cdf $F(x)$. This timer is fine grained, with resolution on the order of inter-frame spacings/backoff timers in 802.11 [8], and causes negligible delay compared to the time for packet transmission. If the timer times out, the receiver transmits a busy tone indicating that it received the packet successfully. Note that the busy tone provides only 1-bit feedback, and there is no information about which receivers have transmitted the busy tone. However, *since every receiver that successfully received the packet independently transmits a busy tone on timer expiry, the time $U$ from the end of packet transmission to the first instant at which the busy tone is turned on is distributed according to the minimum of $X_j$ for all the receivers which successfully received the packet.* Thus, $U$ provides us information to statistically estimate the number of receivers $n$ who successfully got the packet. *In an expected sense, the time delay to the first instant at which the busy tone starts varies inversely as the number of receivers.*

Suppose we know the number of outstanding receivers $x_i$ for packet $i$ before the current transmission. Then, assuming homogeneous channels, the prior probability $q(x_i, n, p)$ that $n$ receivers got the packet is $\binom{x_i}{n} p^n (1-p)^{x_i - n}$. Further, conditioned on $n$ receivers getting the packet, the random variable $U$ has cdf $U_n(x) = 1 - (1 - F(x))^n$ and pdf $u_n(x) = \frac{d}{dx} U_n(x)$. Now suppose the time delay to the first busy tone is measured as $\delta t$. Then, the maximum a posteriori probability (MAP) estimator for this problem is:

$$n^* = \text{argmax}_{n \in \{1,\ldots,x_i\}} q(x_i, n, p) u_n(\delta t).$$

The choice of distribution $F$ is upto the designer, and by choosing it intelligently, one can simplify the computation. For example, if $F(x) = 1 - e^{-cx}$ is exponential with rate $c$, then $U_n(x) = 1 - e^{-ncx}$ is exponential with rate $nc$.

## VII. Related Work

Broadcast disk scheduling [9] emulates a storage device by repeatedly broadcasting packets on a communication channel, focusing on mean [10] and variance [11] of response time. Our work has a different focus, concerned with ensuring the ability to meet real-time constraints while simultaneously exploiting free broadcast capability of wireless communication using techniques from restless bandits in stochastic control. In contrast to static multi-armed bandits, for which the Gittins index policy [3] is optimal, the general restless bandits problem is PSPACE-HARD [4]. This paper applies Lagrangian relaxation techniques for the restless bandits problem [2] to derive a state-dependent index, and derive a heuristic index therefrom for the wireless real-time multicast scheduling problem. This index policy is asymptotically optimal under certain regularity assumptions [12]. Sufficient conditions for Whittle indexability are in general unknown; thus, our result establishing indexability for wireless real-time scheduling is significant. Finally, our work shares the goal of "optimal real-time wireless scheduling" with recent dynamic programming approaches for wireless real-time scheduling in single-rate [13], [14] and multi-rate settings [15].

## VIII. Conclusion

We studied a multicast real-time scheduling problem in a wireless LAN using restless bandits. We established Whittle indexability, and explicitly computed the Whittle index for both the homogeneous and heterogeneous channels under the assumption of complete feedback. We derived performance bounds on how far the Whittle relaxation is from optimality. We designed a noisy feedback protocol to relax the complete feedback assumption. The derived index policy is simple and amenable to low-complexity implementation.

Four challenges remain in the realization of practical wireless multicast real-time scheduling policies. The first is better performance bounds for the index policy. The second is the extension of such performance bounds to incorporate the noisy feedback scheme. The third is in extension of our results to other channel models. The last is in design and evaluation of practical protocols based on these policies.

## References


[1] D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris, "Link-level measurements from an 802.11b mesh network," in *Proceedings of ACM SIGCOMM*. New York, NY, USA: ACM Press, 2004, pp. 121–132.

[2] P. Whittle, "Restless bandits: activity allocation in a changing world," *Journal of Applied Probability*, vol. 25, pp. 287–298, 1988.

[3] J. C. Gittins and D. M. Jones, "A dynamic allocation index for the discounted multiarmed bandit problem," *Biometrika*, pp. 561–565, 1979.

[4] C. H. Papadimitrious and J. Tsitsiklis, "The complexity of optimum queueing network control," *Mathematics of Operations Research*, pp. 293–305, 1999.

[5] V. Raghunathan, "Physical layer aware algorithms and protocols for wireless networks," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2007.

[6] D. Bertsekas, *Dynamic Programming and Optimal Control: Vol I and II*. Athena Scientific, 2005.

[7] C. L. Liu and J. W. Layland, "Scheduling algorithms for multiprogramming in a hard-real-time environment," *Journal of the ACM*, vol. 20, no. 1, pp. 46–61, 1973. [Online]. Available: citeseer.ist.psu.edu/liu73scheduling.html

[8] IEEE 802.11 Wireless Local Area Network Working Group, *IEEE 802.11 Standard*. IEEE Standards for Information Technology, 1999.

[9] Z. Zdonik, R. Alonso, M. Franklin, and S. Acharya, "Are 'disks in the air' just pie in the sky?" in *IEEE Workshop on Mobile Computer Systems*, 1994.

[10] D. Aksoy and M. Franklin, "Rxw: A scheduling approach for large-scale on-demand data broadcast," in *Proceedings of IEEE INFOCOM*, 1998.

[11] S. Jiang and N. H. Vaidya, "Scheduling data broadcast to "impatient" users," in *MobiDE*, 1999, pp. 52–59. [Online]. Available: citeseer.ist.psu.edu/article/jiang99scheduling.html

[12] R. Weber and G. Weiss, "On an index policy for restless bandits," *Journal of Applied Probability*, vol. 27, pp. 637–648, 1990.

[13] V. Raghunathan, M. Cao, and P. R. Kumar, "Real-time scheduling for wireless systems with losses," *IEEE Transactions on Automatic Control (in review)*, 2007.

[14] A. Dua and N. Bambos, "Wireless packet scheduling with deadlines," *IEEE Transactions on Mobile Computing (to appear)*, 2007.

[15] V. Bhandari, V. Raghunathan, B. D. Bui, and M. Caccamo, "Implications of multiple transmission rates on real-time scheduling in wireless networks," in *Proceedings of ACM MOBICOM*, 2007.