Performance guarantees through partial information based control in multichannel wireless networks

Sudipto Guha, Kamesh Munagala, Saswati Sarkar*

September 19, 2006

Abstract

We consider a wireless system with multiple channels when each channel has several transmission states. A user learns about the instantaneous state of an available channel by transmitting a control packet in it. In presence of multiple channels, this process of probing every channel to find the best one consumes significant energy and time. So a user needs to select a channel based only on partial information about instantaneous states of available channels. Furthermore, a user not only needs to optimize its selection based on the information it has, but also needs to determine what and how much information it needs to acquire about the instantaneous states of the available channels. We investigate desired tradeoffs between information acquisition and exploitation in context of wireless networks. We seek to maximize a utility function which depends on both the cost and value of information, and obtain computationally simple joint information acquisition and exploitation strategies that attain constant factor approximations of the optimal solution.

1 Introduction

Future wireless networks will provide each terminal access to a large number of channels. A channel can for example be a frequency in a frequency division multiple access (FDMA) network, or a code in a code division multiple access (CDMA) network, or an antenna or a polarization state (vertical or horizontal) of an antenna in a device with multiple antennas (MIMO). Several existing wireless technologies, e.g., IEEE 802.11a [1], IEEE802.11b [13], IEEE802.11h [2] propose to use multiple frequencies. For example, IEEE 802.11a protocol has 8 channels for indoor use and 4 channels for outdoor use in the 5GHz band, while the IEEE 802.11b protocol has 3 channels in the 2.4 GHz band. The potential deregulation of the wireless spectrum is likely to enable the use of a significantly larger number of frequencies. Due to significant advances in device technology, laptops with multiple antennas (antenna arrays) incorporated in the front lid, and devices with smart antennas have already been developed, and the number of such antennas are likely to significantly increase in near future.

The increase in the number of channels is expected to significantly enhance network capacity and enable several new bandwidth-intensive applications as multiple transmissions can now proceed simultaneously in a vicinity using different channels. Furthermore, the availability of multiple channels substantially enhances the probability (at any given time) of existence of at least one channel with acceptable transmission quality,

*S. Guha is with the department of Computer and Information Science at the University of Pennsylvania. K. Munagala is with the department of Computer Science at Duke University. S. Sarkar is with the department of Electrical and Systems Engineering at the University of Pennsylvania. The authors were supported in part by NSF grants NCR $02-38340$, CCF $04-30376$, CNS $04-543264$ and CNS $05-40347$ and an Alfred P. Sloan Fellowship. Their emails are sudipto@cis.upenn.edu, kamesh@cs.duke.edu, swati@seas.upenn.edu.

since the transmission quality of the individual channels stochastically vary with time and location of the users. These benefits can however be realized only if the users can select the channels efficiently.

Most of the existing channel selection strategies assume complete knowledge of instantaneous transmission qualities of all channels. We refer to this approach as "complete information based optimal control". Note that a user can only learn the instantaneous state of a channel by transmitting a control packet in it and subsequently the receiver informs the sender about the quality of the channel in a response packet (e.g., the RTS and CTS packet exchange in IEEE 802.11). The exchange of control packets in this probing process consumes additional energy, and prevents other neighboring users from simultaneously utilizing the channel. Probing a channel is therefore associated with a cost. When the number of available channels is large, the cost incurred in learning the instantaneous transmission qualities of all channels may become prohibitive. Owing to this cost, some recent papers have investigated selection strategies that assume no knowledge of instantaneous transmission qualities of any channel [25]. This approach, which we refer to as "minimal information based optimal control", may however attain significantly lower transmission rates owing to sub-optimal selection of channels. We seek to design a framework that attains any desired tradeoff between the above extremes using only simple control mechanisms. Specifically, we develop a framework for *partial information based stochastic control* which, in accordance with the cost and the benefits of probing different channels, determines both (a) the amount of information a user must obtain about the instantaneous transmission qualities of the channels at its disposal and also (b) how to select the channels based on the acquired information.

We consider a single sender with access to n channels. The instantaneous transmission qualities of the channels can have K possible values and stochastically vary with time. The statistics of these temporal variations may be different for different channels. Every time the sender probes a channel it learns about the signal to noise ratio and thereby the probability of success in the channel, but also incurs a certain cost which may again be different for different channels. Before each transmission, the sender needs to determine how many and which channels it will probe and also the sequence in which these channels will be probed (*probing policy*). Note that depending on the available hardware (e.g., availability, or lack thereof, of multiple network interface cards, or compatible transmission circuits to appropriately distribute the power across the antennas), a sender may, or may not, be able to simultaneously transmit in multiple channels. In this paper, we consider the scenario where a sender can transmit in only one channel in a time slot and transmits at most one packet in each slot. Based on the outcomes of the probes, a sender decides whether to transmit or defer transmission until transmission qualities improve (*transmission policy*). If the sender decides to transmit, it must select one of the available channels (*channel selection policy*), which need not be those that it has probed.

The sender seeks to maximize a utility function which depends on the probability of success of each transmission and the probing cost incurred before each transmission. Clearly, a meaningful utility function should (a) increase with increase in the former and (b) decrease with increase in the latter. As a starting point, we consider linear functions that satisfy the above criteria. Specifically, we seek to design a jointly optimal probing and channel selection policy that maximizes a system utility which is the difference between the probability of successful transmission and a suitably scaled expected probing cost before each transmission. Loosely, this utility function represents the gain or the profit of the sender if the sender receives credit from the receiver for each packet it delivers successfully and needs to additionally compensate the wireless provider for each probe packet it transmits[1].

Different choices of the scaling factor of the probing cost in the utility function lead to different intermediate points between the extremes of complete information based optimal control and minimal information based optimal control – a small scaling factor corresponds to complete information (or probing everything),

[1]The sender may have to share with the provider part of the credit it receives from the receiver for each successfully delivered packet. Then the credit we are considering here is the credit remaining after the sharing process.

while a large scaling factor corresponds to minimal information (or absence of probing).

We first enumerate the challenges in designing the jointly optimal strategy. The optimal policy needs to probe adaptively, i.e., the result of a probe determines the channels to be probed subsequently. For example, consider channels with 3 possible states $(0, 1, 2)$, each of which is associated with a different transmission quality. Clearly, the probing terminates if a probed channel is in the highest state. Now, let a probed channel be in the intermediate state (state 1). Then the subsequent probes should be limited to channels that have high probabilities of being in the highest state. However, if all channels that have been probed in a slot are in the lowest state, then the channels that have high probabilities of being in the intermediate state may also be subsequently probed. Also, the channel selection decision depends on the outcomes of the probes and the expectation and uncertainty of the transmission quality of the channels that have not been probed. The optimal policy is therefore a decision tree over n variables (Figure 1) – naive computations will require both exponential time and exponential storage space. Next, the policies may depend on the higher order statistics of the channels. This is because the optimum policy may not probe a channel if its quality has a low variance as probing it does not provide significant information but incurs additional cost.

Contributions: Our main contribution is to obtain succinct polynomial time computable joint probing, selection and transmission policies that provably attain utility values which are within constant factors of the optimal utilities. When a sender always has packets to transmit (saturated sender assumption), this constant factor is $4/5$ (Section 4); otherwise the constant factor is $2/3$ (Section 5). The policies can be readily implemented in resource constrained devices as their execution times and storage spaces grow only linearly with increase in input size. Our results are somewhat surprising given that optimal solutions for most partial information based control problems turn out to be computationally intractable, and standard approximation techniques either do not provide guaranteed approximation ratios or require exponential computation times [5]. Our proofs therefore rely on exploitation of specific system characteristics and employ techniques that are not standard in context of stochastic control. Finally, the area of partial information based control problems, and in particular the joint optimization of the reward obtained from informed selections and the cost incurred in acquiring the required information, remains largely unexplored in wireless networks. Thus, both our results and proofs will enhance the state of knowledge in an emerging area which has hitherto received only limited attention.

Our assumptions are general in that we assume arbitrary probing costs and distributions of the transmission qualities of the channels and also arbitrary number of channels and transmission states for the channels (Section 3). The only restrictive assumption we make is that we assume that the states of channels are temporally and spatially uncorrelated (earlier results in this area also rely on the same assumption [18, 23]). Given the large number of possible courses of actions, obtaining policies with provable performance guarantees is non-trivial even under this assumption, and is probably a pre-requisite for exploring more general scenarios involving temporally correlated channels and networks with multiple terminals. Our current policies directly lead to heuristics in these cases, and our empirical investigations indicate that these heuristics attain near-optimal performance. We hope that our work will provide the framework for obtaining provable performance guarantees in these complex scenarios.

2 Related Literature

We first discuss the relation of our problem with some classical problems like the stopping time and multi-armed bandit problems. The most well-researched version of the stopping time problem is a stochastic control problem that optimally selects between two possible actions at any given time: to continue or to stop [8]. Recently, the results for this problem have been used to solve partial information based control problems for statistically identical channels with equal probing costs [18, 23]. Since we consider channels that may have different statistics and/or different probing costs, the optimal action needs to be selected from multiple options at any given time - the options being (a) whether to continue probing (b) which channel to probe next if the decision is to probe and (c) which channel to transmit if the decision is to stop probing. Thus, the results from the above version of stopping time problem do not apply in our context. The optimal stopping time problem has also been considered in a more general setting where the number of available actions may be more than two; our problem is in fact a special case of this general version (Chapter IV, [5]). In this general case, the process terminates in certain states, which constitute the termination set, and selects the optimal action in other states. But, so far, only certain broad characterizations of the termination set are known in this general case, and the optimal actions when the decision is not to stop are also not known in close form [5]. Thus, these general results do not lead to the optimal policies we are seeking to characterize.

The stochastic multi-armed bandit problem considers a bandit with n arms [12]. The system can try one arm in each slot, and when it tries an arm, it receives a random reward which depends on the state of the arm. The state of an arm changes only when the system tries it. The reward of a system in T slots is the sum of the rewards in each slot. The goal is to maximize the expected reward in T slots. Our problem differs from the above in that (a) the state of a channel can change even when it is not probed or used for transmission and (b) a node can learn the states of multiple channels in an epoch while incurring additional probing costs for learning the state of each additional channel. The adversarial multi-armed bandit problem removes one of the above differences in that it allows the state of an arm to change even when the system does not try it [3]. But, it seeks to optimize the selection under the assumption that the sender uses the same arm in all slots. Note that we allow a sender to probe and transmit in different channels in different slots. In another version of the adversarial multi-armed bandit problem, the goal is to select the arms so as to minimize the "regret" or the difference in expected reward with the best policy in a collection of a certain number (say N) of given policies. As expected, the regret in T slots increases with increase in both N and T (the regret for the best known policy is $O(\sqrt{nT\ln(N)})$). In our context, the total number of possible probing and channel selection policies that can be used in T slots is large, e.g., the number of deterministic policies is $(n!)^T$. Thus the results available in this context do not apply in our problem, and we use different solution approach and obtain different performance guarantees.

Optimizing the order of evaluation of random variables so as to minimize the cost of evaluation ("pipelined filters") has been investigated in several different contexts like diagnostic tests in fault detection and medical diagnosis, optimizing conjunctive query and joint ordering in data-stream systems, web services [4, 6, 9, 11, 17, 19, 20, 21, 22]. However our work is different from all the above in that, we (a) consider multi-state channel models whereas pipeline filters consider two state models and (b) allow a node to transmit in a channel even if the channel has not been probed. Note that usually two state models can not capture the statistical variations of wireless channels [14]. As we demonstrate later, both the above generalizations significantly alter the decision issues and the optimal solutions (Section 3).

Finally, opportunistic selection of channels with complete knowledge of channel states has been comprehensively investigated over the last decade (e.g., [24]). But, in general, the area of partial information based control problems, and in particular the joint optimization of the reward obtained from informed selections and the cost incurred in acquiring the required information, remains largely unexplored in wireless networks. The first results in this area have been obtained in [18, 23], but as mentioned above, they consider only statistically identical channels with equal probing costs. Recent statistical investigations indicate that different channels available to a sender may have different statistics [14]. We now describe our initial results in the area. We have recently proved that when every channel has two states the joint optimization problem can be solved in polynomial time even when different channels have different statistics and probing costs [16]. In [15], we proved that for channels with multiple states the optimization can be approximated within a factor of $1/2$ using polynomial time algorithms. In the current paper, we significantly enhance the results obtained in [15]. Specifically, we prove that the joint optimization problem can be approximated within a factor of $4/5$ for channels which have different statistical distributions, different probing costs and multiple states. We also obtain analytical guarantees under significant generalizations of the framework. Specifically, we relax the assumptions that (a) the sender always has a packet to transmit and (b) the channel states are temporally independent. The results in this paper will therefore enhance the state of knowledge in an emerging area which has hitherto received only limited attention.

3 System Model and Problem Definition

A sender U has access to n channels which are denoted as channels $1, 2, \ldots, n$, each of which has K possible states, $0, \ldots, K-1$. We assume that time is slotted. In any slot channel j is in state i with probability p_{ij} independent of its state in other time slots and the states of other channels in any slot. Without loss of generality, we assume that $p_{K-1j} < 1$ for each j, as otherwise the optimum policy is simply to transmit in j without probing any channel. In every slot, U transmits at most one data packet in a selected channel. If the channel selected for transmission is in state i, the transmission is successful with probability r_i. Without loss of generality we assume $0 \leq r_0 < r_1 < \cdots < r_{K-1}$. For simplicity, we also assume that $r_0 = 0$; all analytical results can however be generalized to the scenario where $r_0 > 0$. Whenever U probes a channel j, it pays a cost of $c_j \geq 0$. Probing different channels may incur different costs as the probing process for different channels may interfere with the channel access of different number of users (based on geometry and allocation of channels). We now formally define the policies and the performance metrics.

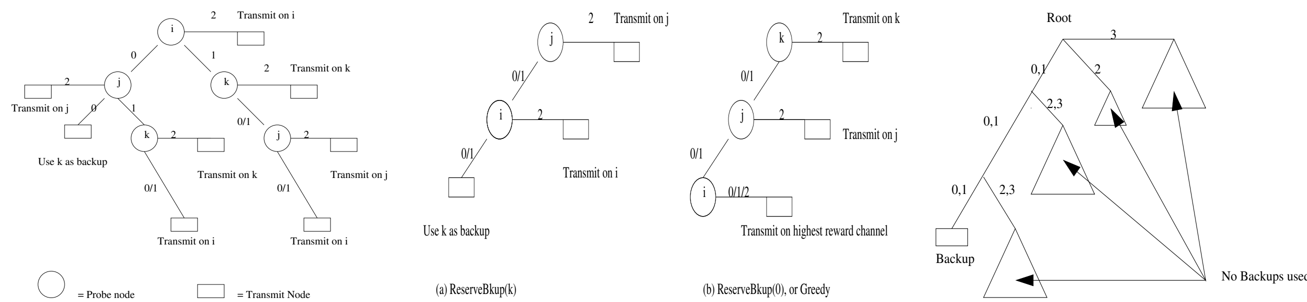
Definition 3.1 *A* **probing policy** *is a rule that, given the set of channels the sender has already probed in a slot (which would be empty at the beginning of the slot) and the states of the channels probed in the slot, determines (a) whether the sender should probe any more channels and (b) if the sender probes additional channels which channel it should probe next. The sender knows the state of a channel in a slot if and only if it probes the channel in the slot.*

Definition 3.2 *A* **selection policy** *is a rule that selects a channel for the transmission of a data packet in a slot on the basis of the states of the probed channels, after the completion of the probing process in the slot. The selection policy can select a channel even if it has not been probed in the slot, and in that case, the channel is referred to as a* **backup** *channel.*

Definition 3.3 *The* **probing cost** *is the sum of the costs of all channels probed in the slot. The probing cost is clearly a random variable that depends on the probing policy and the outcomes of the probes (as the sender may probe subsequent channels depending on the outcomes of the previous probes). The* **expected probing cost** *is the expectation of this random variable and depends on both the probing policy and the channel statistics.*

Definition 3.4 *In any slot, the* **transmission reward** *is* 1 *if there is a successful transmission and* 0 *otherwise. Therefore, the expected transmission reward is* r_i *in a slot* t *if* U *transmits in a channel in state* i *during* t*. We overload the terminology and denote the "reward" of transmitting in state* i *by* r_i*. The expected transmission reward of a policy is therefore* $\sum_i q_i r_i$ *where* q_i *is the probability that the selection policy decides to use a channel which in state* i*;* q_i *depends on the channel statistics as well as the policy.*

Definition 3.5 *The* **expected utility** *of the sender, denoted simply as* **gain***, is the difference between the expected transmission reward, and the probing cost scaled by a factor* κ*. The gain depends on the probing and selection policies, the channel statistics and the scaling parameter* κ *– choosing the scaling parameter* κ *to be* 0 *makes the policy acquire complete information, while setting it to* ∞ *makes the policy acquire no information. Since* κ *can be included in the probing costs themselves, we drop this parameter in the remaining discussion without loss of generality.*



(a) OPT (b) RESERVEBKUP (c) Proof of lemma 4.1

Figure 1: Figure (a) shows the decision tree for OPT in Example 2.1. A channel is probed at each probe node, and the letter inside it indicates which channel is probed at the node. The numbers next to the branches indicate the outcome of the probe. The number r/s next to a branch indicates that both states r and s of the previously probed channel lead to the same action. For example, the sender first probes channel i. If i is in state 2, it transmits in i. If i is in state 1 and 0, it probes k and j respectively. The first and second figure in Figure (b) respectively illustrates RESERVEBKUP(k) and RESERVEBKUP(0) for the scenario described in Example 2.1. Note that consistent with the definition RESERVEBKUP(0) never uses a backup channel. Also, RESERVEBKUP(k) attains the maximum gain in $\cup_{\ell \in \{0,i,j,k\}} \mathcal{P}(\ell)$, but this maximum gain is lower than that of OPT. Figure (c) illustrates the transformation of OPT2 to OPT1 in lemma 4.1.

In Section 4, we assume that U's queue is never empty (saturated sender assumption); we relax this assumption in Section 5. The two versions of the problem are defined as follows:

Saturated Sender Problem: Under the saturated sender assumption, at least one policy that maximizes the utility transmits a packet in every slot. We therefore assume that U transmits exactly one data packet in every slot. The problem formulation for the saturated sender case follows.

Problem 1 *Given $\{c_j\}, \{r_i\}$ and $\{p_{ij}\}$, find a probing and selection policy so as to maximize the expected gain. Let* OPT *denote the optimal policy and G_{OPT} its gain.*

Unsaturated Sender Problem: We now relax the assumption that a sender always has packets to transmit. A sender generates packets as per an arrival process which constitutes a positive recurrent, aperiodic, irreducible Markov chain. Under the steady state distribution of the Markov chain, the expected number of arrivals in any slot is λ, where $\lambda \in [0, 1)$. Packets are stored in an infinite buffer. If in a slot the sender has a packet in its queue, the slot is referred to as a *busy* slot. The sender transmits only in busy slots, but may not transmit in every busy slot; it may improve its gain by deferring transmission until at least one channel has good quality. The *transmission policy* is a rule that determines which slots a sender transmits. The decisions may depend on the outcomes of the probes, the queue lengths, channel and arrival statistics. The sender must however ensure that it transmits at least at the rate at which it generates packets, otherwise its delay becomes unbounded. In addition to gain, system stability is therefore of interest.

Definition 3.6 *The system is stable if the sender's expected queue length is finite. A policy that attains finite expected queue length is a stable policy.*

Problem 2 *Given $\{c_i\}, \{r_i\}$ and $\{p_{ij}\}$ find a probing, selection and transmission policy that stabilizes the system and maximizes the gain among all stable policies. Let* UNSAT *denote the optimal policy and G_{UnSat} its gain.*

Policies and Decision Trees: Every joint probing and selection policy can be represented by a unique decision tree (Figure 1); we therefore use policies and decision trees interchangeably.

3.1 Challenges

Let us first consider the saturated sender case. The optimal probing policy does not probe any further in a slot if a probed channel is in state $K-1$. Since channels are temporally independent, the optimal probing and selection strategies in a slot need not depend on the decisions and the observations in other slots. Also, the optimal probing and selection strategies remain the same in all slots, though the specific choices made by each policy may be different in different slots depending on the outcome of the probes. Using these observations, the optimal policy can be computed using a bottom-up dynamic program. However, the computation time for the optimal is $\Omega(K2^n)$ time as the dynamic program has $K2^n$ states, and the storage space is $\Omega(n^K)$.

Minor perturbations in the parameter values and policy constraints substantially alter the structure of the optimal policy. This suggests that significant reduction in computation time and storage space may not be possible for the optimal policy. For example, when $K = 2$ (i.e., each channel is either "on" or "off") then the optimal policy probes channels in increasing order of c_j/p_{2j} (i.e., the ratio between their costs and their probabilities of being in the "on" state) [16]. For $K > 2$, the natural generalizations of this policy are to probe channels in increasing order of the ratio between their (a) costs and the probabilities of being in the highest state (i.e., c_j/p_{K-1j}) or (b) costs and the expected rewards (i.e., $c_j/\sum_{K=1}^{K} p_{kj}r_j$). We next present one example where both these probing sequences are sub-optimal.

Example 3.1 *(Saturated Sender Case.) Let U have access to 3 channels i, j, k each of which has 3 states. Let $r_2 = 1, r_1 = 0.1$. The probabilities associated with different states of i, j, k are $(0.49, 0.02, 0.49)$, $(0.5, 0.01, 0.49)$, $(0.5, 0.5-\delta, \delta)$ respectively. Also, $c_i = 0.05885\delta, c_j = 0.06\delta, c_k = 0.05\delta$. Let $\delta < 0.15$. Note that i has the least, j intermediate and k maximum expected rewards. Also, $c_k/p_{2k} < c_j/p_{2j} < c_i/p_{2i}$. Thus, intuitively, if i and j are both probed, then j should be probed before i. But,* OPT *probes i before probing j (Figure 1(a))!*

Note that if the transmission policy is constrained to transmit only in a probed channel (i.e., a backup channel is not allowed), then the optimal probing policy in fact probes channels in sequence of c_j/p_{K-1j} for arbitrary K [15]. But, the previous example demonstrates that the availability of a backup channel renders this probing sequence suboptimal.

The situation is significantly more complicated in the unsaturated sender case – because in addition to a backup, the sender now has a choice of **not transmitting at all**, that is, if she believes that she has been unlucky in the current time slot and wishes to try to send the packet in subsequent slots. We note that no previous results – even exponential time algorithms, were known for this problem.

4 The Saturated Sender Problem

We first establish a key technical property for the optimal policy which states that whenever it transmits in a channel that has not been probed, it transmits in a fixed channel whose selection does not depend on the outcome of the probes (Subsection 4.1). We refer to this result as the *Structure Theorem* (Theorem 4.2). Using the Structure Theorem, we design a computationally simple probing and selection policy that attains at least $4/5$ of the maximum gain irrespective of $n, K, \{p_{ij}\}, c_j$ (Subsection 4.2).

4.1 Structure Theorem for the optimal policy

We first introduce the following notion which will be useful in stating and proving the Structure Theorem.

Definition 1 *A $\leq i$ tree is a decision tree which takes the same decisions irrespective of the states of the probed channels provided the states are less than or equal to i. The decisions corresponding to the states which are less than or equal to i therefore constitute a path in such a tree which we refer to as a $\leq i$ path.*

For example, both decision trees in Figure 1(b) are ≤ 1 trees, and the left-most paths are the ≤ 1 paths. We prove the Structure Theorem 4.2 using the following lemma.

Lemma 4.1 *Suppose* OPT *probes a channel j at a node m in its decision tree and if j is in state i it takes a backup downstream. Then there exists another optimum policy* OPT1 *which has the same decision tree as* OPT *except possibly for the tree rooted at m. In* OPT1*, the tree rooted at m,*

1. *is a $\leq i$ tree*
2. *takes a backup, say ℓ, at the end of its $\leq i$ path and*
3. *takes ℓ as a backup wherever it takes a backup.*

Proof: We prove the lemma using induction on the states. The lemma holds by vacuity for all channels j, node m if $i = K - 1$. This is because OPT never takes a backup after observing a channel in state $K - 1$.

Suppose the lemma holds for all channels j, nodes m in the decision tree of OPT, states $i+1, \ldots, K-1$. We prove the lemma for all channels j, nodes m and state i. Now, let OPT probe a channel j at a node m in its decision tree and take a backup somewhere downstream if j is in state i. Let m_1 be the first node after j is probed at m and observed to be in state i. Clearly, the decision tree rooted at m_1 is a $\leq i$ *tree*, and takes at least one backup somewhere downstream.

We will first show that there is one optimum policy OPT2 which is similar to OPT except possibly for the tree rooted at m_1. The tree rooted at m_1 in OPT2 is still a $\leq i$ *tree* but

- p1 takes a backup, say ℓ, at the end of its $\leq i$ *path* and
- p2 takes ℓ as a backup wherever it takes a backup.

Suppose the tree T rooted at m_1 in OPT does not satisfy the above conditions. Then there is a path originating from its $\leq i$ *path* which ends in a backup. Let m_2 be the highest node (i.e., node closest to m_1) from where such a path originates on the $\leq i$ *path* of T. Clearly this path corresponds to a channel being observed in a state higher than i, say q, at m_2. From the induction hypothesis, there exists one optimum OPT2 which is similar to OPT except possibly for the decision tree rooted at m_2. In OPT2 this decision tree

- is a $\leq q$ *tree*
- takes a backup, say ℓ, at the end of its $\leq q$ *path* and
- takes q as a backup wherever it takes a backup.

Note that OPT2 satisfies conditions **p1** and **p2** for the tree rooted at m_1 (refer Figure 1(c)).

Let α be the probability that OPT2 never visits m, G' be the expected gain of OPT2 if it never visits m, G_h be the expected gain of OPT2 given that j is observed in state h at node m. Thus, the expected gain of OPT2 is $\alpha G' + (1 - \alpha) \sum_h p_{hj} G_h$. Let T_h be the decision tree in OPT2, and hence in OPT, after j is observed to be in a state $h < i$ after being probed at m. Consider a new policy which is similar to OPT2 except that it replaces the decision tree rooted at m_1 with T_f for some $f < i$. The expected gain of this new policy is at least $\alpha G' + (1-\alpha)\sum_{h,h\neq i} p_{hj} G_h + (1-\alpha)p_{ij}G_f$. Since the expected gain of this policy can not exceed that of the optimum, $G_f \leq G_i$.

Now, consider another policy OPT1 which is similar to OPT2 except that it replaces the decision trees $T_0, \ldots, T_{i-1}$ (i.e., those rooted at nodes immediately downstream of m and corresponding to j being in states lower than i), with the decision tree rooted at m_1 (i.e., the one corresponding to j being in state i) (refer Figure 1(c)). Since the decision tree rooted at m_1 is a $\leq i$ tree and the $\leq i$ path ends in a backup, the gain of OPT1 is $\alpha G' + (1-\alpha)\left(G_i \sum_{h\leq i} p_{hj} + \sum_{h>i} p_{hj} G_h\right)$. Thus, since $G_f \leq G_i$, for $f < i$, the expected gain of OPT1 is at least as high as that for OPT2. Thus, OPT1 is also optimum. Note that OPT1 is similar to OPT except possibly for the decision tree rooted at m, which is a $\leq i$ *tree* and satisfies conditions **p1** and **p2**. The result follows. □

We now state and prove the Structure Theorem.

Theorem 4.2 (Structure Theorem) *There exists an optimum policy that uses a unique backup channel. The backup channel is used at the end of one $\leq i$ path.*

Proof: Consider an optimum policy that does not use a backup channel at all. Consider the path in its decision tree which corresponds to all probed channels being in state 0. Modify the policy so as to use the last channel probed in this path as a backup. Note that the gain does not decrease. Thus, the modified policy is also optimum. Thus, there always exists an optimum policy that uses a backup channel at the end of some path in its decision tree. Let one such optimum policy OPT3 use backup channels at the end of at least two distinct paths P_1 and P_2 in the decision tree. Let P_1 and P_2 intersect at node m in the decision tree. Then, OPT3 probes a channel j at m and uses P_1 and P_2 for two different states of j. Let k be the higher of the two states. Then, by lemma 4.1, there exists another optimum policy OPT4 which differs from OPT3 only in the decision tree rooted at m which is a $\leq k$ tree in OPT4. Clearly, OPT4 has strictly fewer number of paths ending in backups. The Theorem follows from recursive usage of this argument. □

Let $\Theta(\ell)$ be the class of policies that never use any channel other than ℓ as a backup. It follows from the Structure Theorem that OPT is in $\Theta(\ell)$ for some ℓ. Thus, if the optimum in $\Theta(\ell)$ can be computed in polynomial time for each ℓ then the overall optimum can also be computed in polynomial time. The difficulty in computing the optimum in any given $\Theta(\ell)$ is in determining the optimum probing sequence and also in deciding whether to probe ℓ or use it as a backup. Both these decisions may depend on the outcomes of the probes. In the next subsection, we prove that there exists ℓ such that the policy that is optimum in a subset of $\Theta(\ell)$, $\mathcal{P}(\ell)$, attains $4/5$ of the overall maximum gain, G_{OPT}. We also show that for each ℓ the optimum in $\mathcal{P}(\ell)$ can be computed using a simple polynomial time algorithm, ReserveBkup(ℓ). Thus $4/5$ of G_{OPT} can be attained in polynomial time as well.

4.2 A $4/5$ approximation algorithm

Definition 4.1 *Let $\mathcal{P}(\ell)$ denote the class of policies, each of which (a) never probes ℓ and (b) never uses any channel other than ℓ as a backup. The policy that maximizes the gain among all policies in $\mathcal{P}(\ell)$ is denoted as* OPT*(ℓ). Let G_{ALG} be the maximum gain attained by some policy in $\cup_{l=0}^{n}\mathcal{P}(\ell)$.*

Since the channels are numbered $1, \ldots, n$, by definition, $\mathcal{P}(0)$ corresponds to the class of policies that never use backup channels.

The following lemma indicates why the above class of policies are of interest.

Lemma 4.3 *Let ℓ be the unique backup associated with* OPT. *Let* OPT *also probe ℓ with probability $1-\alpha$. Consider a policy B that is similar to* OPT *except that it never probes ℓ; instead wherever* OPT *probes ℓ,*

B follows the same course of actions as OPT *does after discovering ℓ in state* 0. *Then B attains a gain of at least $G_{OPT} - (1-\alpha)(\sum_{i=0}^{K-1} p_{i\ell} r_i - c_\ell)$.*

Proof: Let OPT probe ℓ at nodes numbered $1, \ldots, J$ in its decision tree. Let $\beta_1, \ldots, \beta_J$ be the respective probabilities that OPT traverse these nodes and $G_1, \ldots, G_J$ be the respective gains of OPT given that it traverses these nodes. Note that the decision tree of B is the same as that for OPT except for the trees rooted at $m_1, \ldots, m_J$. B reaches these nodes with the same probabilities as OPT. Let $G'_1, \ldots, G'_J$ be the gains of B given that it traverses these nodes. Then the difference between the overall gains of OPT and B is $\sum_{k=1}^{J} \beta_k (G_k - G'_k)$. We will show that for each k, $G_k - G'_k \leq \sum_{i=0}^{K-1} p_{i\ell} r_i - c_\ell$. The result follows since $\sum_{k=1}^{J} \beta_i = 1 - \alpha$.

Consider m such that $1 \leq m \leq J$. Let q be the highest state of a probed channel before OPT (and hence B) reaches node m in their decision trees. Let T_i be the decision tree in OPT after ℓ is observed to be in a state i after being probed at m. Let C_i be the expected probing cost in T_i, $\mathcal{A}_i$ be the set of paths from the root to the leaves of T_i, γ_P be the probability that OPT traverses along path P given that it traverses T_i, and ν_P be the highest state of a channel probed in path P. Note that downstream of m, OPT transmits only in probed channels. Thus,

$$
\begin{aligned}
G_k &= \sum_{i=0}^{K-1} p_{i\ell} \left(\left(\sum_{P \in \mathcal{A}_i} \gamma_P \max(r_q, r_i, \nu_P) \right) - C_i \right) - c_\ell \\
&= \sum_{i=0}^{K-1} p_{i\ell} \left(\max(r_q, r_i) - C_i + \sum_{\substack{P \in \mathcal{A}_i \\ \nu_P > \max(r_q, r_i)}} \gamma_P \left(\nu_P - \max(r_q, r_i) \right) \right) - c_\ell \\
&= r_q \sum_{i=0}^{q} p_{i\ell} + \sum_{i=q+1}^{K-1} p_{i\ell} r_i - c_\ell + \sum_i p_{i\ell} \tau_i,
\end{aligned}
$$

$$
\text{where } \tau_i = \sum_{\substack{P \in \mathcal{A}_i \\ \nu_P > \max(r_q, r_i)}} \gamma_P \left(\nu_P - \max(r_q, r_i) \right) - C_i.
$$

Now, consider another modification of OPT that uses T_i in place of T_0 in the decision tree of OPT. The gain of this policy given that it traverses node m is $r_q \sum_{i=0}^{q} p_{i\ell} + \sum_{i=q+1}^{K-1} p_{i\ell} r_i - c_\ell + \sum_{i=1}^{K-1} p_{i\ell} \tau_i + p_{0\ell} \theta_i$ where $\theta_i = \sum_{\substack{P \in \mathcal{A}_i \\ \nu_P > r_q}} \gamma_P (\nu_P - r_q) - C_i$. This gain must be upper bounded by G_k from the optimality of OPT. Thus, $\tau_0 \geq \theta_i$. Note that $\theta_i \geq \tau_i$. Thus, $\tau_0 \geq \tau_i$ for each i. Thus,

$$
G_k \leq r_q \sum_{i=0}^{q} p_{i\ell} + \sum_{i=q+1}^{K-1} p_{i\ell} r_i - c_\ell + \tau_0. \tag{1}
$$

Now consider B.

$$
\begin{aligned}
G'_k &= \sum_{P \in \mathcal{A}_0} \gamma_P \max(r_q, \nu_P) - C_0 \\
&= r_q + \tau_0
\end{aligned} \tag{2}
$$

From (1) and (2), $G_k - G'_k \leq \sum_{i=0}^{K-1} p_{i\ell} r_i - c_\ell$. □

We now propose a policy RESERVEBKUP(l) that attains the maximum gain in $\mathcal{P}(l)$ for $l = 0, \ldots, n$.

Definition 4.2 *For $i = 1, \ldots, n$,, let $\tilde{r}_i[u] = \frac{\sum_{v=u}^{v=K-1} p_{vi} r_{vi}}{\sum_{v=u}^{v=K-1} p_{vi}}$ and $\tilde{p}_i[u] = \sum_{v=u}^{v=K-1} p_{vi}$. Let $\tilde{r}_0[0] = -1$ and $r_{-1} = -1$. Let $w_\ell = \min_u \{u : r_u > \tilde{r}_\ell[0]\}$*[2].

Definition 4.3 *Let $H_{u,\ell} = \phi$ for all $u \geq K$. For each ℓ, starting from $u = K-1$, down to $u = w_\ell$, recursively, define $H_{u,\ell} = \left\{ i | i \notin \bigcup_{v:v>u} H_{v,\ell}, \text{ and } \tilde{r}_i[u] - \frac{c_i}{\tilde{p}_i[u]} > \max(\tilde{r}_\ell[0], r_{u-1}) \right\} \setminus \{\ell\}$. Assume $c_i/\tilde{p}_i[u] = +\infty$ when $\tilde{p}_i[u] = 0$.*

RESERVEBKUP(ℓ)

1. Consider each $H_{u,\ell}$ in decreasing order of u starting from $u = K-1$ down to $u = w_\ell$.
2. (**Probing Process:**) Within each $H_{u,\ell}$ probe in non-increasing order of $\tilde{r}_j[u] - \frac{c_j}{\tilde{p}_j[u]}$, and stop if any channel is found to be in state u or above.
3. (**Selection Process:**) Consider the channel, say j, which is in the highest state, say y, among all probed channels. If no channel is probed, $j = -1$. If $r_y \geq \tilde{r}_\ell[0]$, transmit in y, else transmit in ℓ.

Refer to Figure 1(b) for examples elucidating RESERVEBKUP(ℓ). We now explain the design of RESERVEBKUP(ℓ).

First, observe that RESERVEBKUP(ℓ) $\in \mathcal{P}(\ell)$. This is because by definition $\ell \notin H_{u,\ell}$ for any u. Thus, RESERVEBKUP(ℓ) never probes ℓ (refer to the probing process). Also, note that RESERVEBKUP(ℓ) does not use any channel other than ℓ as backup (refer to the selection process).

Next, the channel selection for RESERVEBKUP(ℓ) is clearly optimal among policies in $\mathcal{P}(\ell)$. Note that $\tilde{r}_\ell[0]$ is the probability of success when the sender transmits in the backup ℓ. Thus, if the highest state of a probed channel is u when the optimum policy in $\mathcal{P}(\ell)$, OPT(ℓ), terminates its probing process, then OPT(ℓ) transmits in the probed channel if $r_u \geq \tilde{r}_\ell[0]$ and transmits in ℓ otherwise. This is exactly what RESERVEBKUP(ℓ) does as well.

We now explain the intuition behind the design of the probing process for RESERVEBKUP(ℓ). First let $\ell = 0$. Once a sender observes that a probed channel is in state u it can not increase its gain any further by discovering another probed channel in state u or in a lower state. Thus, subsequently, it probes only the channels j for which the additional reward ($\tilde{r}_j[u+1]\tilde{p}_j[u+1] - r_u \tilde{p}_j[u+1]$) exceeds the cost c_j, i.e., it probes the channels in $H_{v,\ell}$, $v > u$. The incremental gain for probing a channel j then is $\tilde{r}_j[u+1]\tilde{p}_j[u+1] - r_u \tilde{p}_j[u+1] - c_j$.The probing sequence in each $H_{u,\ell}$ follows an increasing order of the ratio between this incremental gain and the probability that the channel is in a state that is higher than the highest observed state u ($\tilde{p}_j[u+1]$). We now comment on the major differences between the probing processes of RESERVEBKUP(ℓ) for $\ell > 0$ and $\ell = 0$. Note that $\tilde{r}_\ell[0]$ is the gain if the sender transmits in ℓ without probing any channel j. If, the sender observes a channel to be in state u for which $r_u \leq \tilde{r}_\ell[0]$, the observation does not increase the gain as compared to $\tilde{r}_\ell[0]$. Hence, the sender considers the incremental gain as $\tilde{r}_j[u+1]\tilde{p}_j[u+1] - \max(r_u, \tilde{r}_\ell[0])\tilde{p}_j[u+1]$ instead of $\tilde{r}_j[u+1]\tilde{p}_j[u+1] - r_u\tilde{p}_j[u+1]$, and, as before, probes only channels for which the incremental gain exceeds the probing cost.

Theorem 4.4 RESERVEBKUP(ℓ) *attains the maximum gain among all policies in $P(\ell)$.*

The proof of the above theorem is given in Subsection 4.3. The gain of RESERVEBKUP(ℓ) can be computed in $O(nK)$ time for each ℓ. Thus, the RESERVEBKUP(ℓ) with the maximum gain can be determined in $O(n^2K)$ time. Once computed, RESERVEBKUP(ℓ) can be executed and stored in $O(n)$ time and memory respectively for any ℓ. We can now prove the main theorem of this subsection.

[2] Note that w_ℓ is well-defined for each ℓ as (a) $\tilde{r}_\ell[0] \geq r_0 = 0$ and (b) $\tilde{r}_\ell[0] < r_{K-1}$ which follows since $p_{K-1\ell} < 1$ for each ℓ.

Theorem 4.5 *The maximum gain of of our algorithm G_{ALG} is at least $4/5 G_{OPT}$.*

Proof: By the structure Theorem (Theorem 4.2), there exists an optimum policy OPT that uses a unique backup, say ℓ. Let α denote the probability with which OPT uses ℓ as a backup. Construct a new policy A that is similar to OPT except that it probes ℓ whenever OPT uses ℓ as a backup. Clearly, A attains a gain of at least $G_{\text{OPT}} - \alpha c_\ell$. Also, since $A \in \mathcal{P}(0)$, its gain is at most G_{ALG}. Thus,

$$G_{\text{ALG}} \geq G_{\text{OPT}} - \alpha c_\ell. \tag{3}$$

Next, consider policy B as described in lemma 4.3. From lemma 4.3, B attains a gain of at least $G_{\text{OPT}} - (1-\alpha)(\sum_{i=0}^{K-1} p_{il} r_i - c_\ell)$. Again, since $B \in \mathcal{P}(\ell)$,

$$G_{\text{ALG}} \geq G_{\text{OPT}} - (1-\alpha)(\sum_{i=0}^{K-1} p_{il} r_i - c_\ell). \tag{4}$$

From multiplying Equations (3) and (4) with $(1-\alpha)$ and α respectively, and adding the results, we have $G_{\text{ALG}} \geq G_{\text{OPT}} - \alpha(1-\alpha)\sum_{i=0}^{K-1} p_{il} r_i$. Now, the policy that uses ℓ as a backup without probing any channel attains a gain of $\sum_{i=0}^{K-1} p_{il} r_i$. Since this policy is in $\mathcal{P}(\ell)$, $\sum_{i=0}^{K-1} p_{il} r_i \leq G_{\text{ALG}}$. Thus, $G_{\text{ALG}} \geq G_{\text{OPT}}/(1+\alpha(1-\alpha))$. The result follows since the maximum value of the denominator is 1.25. □

Finally, in practice, the gain of RESERVEBKUP(ℓ) substantially exceeds the lower bound in Theorem 4.5. In fact, RESERVEBKUP(ℓ) attained the optimal gain for some ℓ in all but one case we considered. For instance, in Example 3.1 (Figure 1(a),(b)), RESERVEBKUP(ℓ) turns out to be suboptimal for all ℓ. The suboptimality arises due to the fact that RESERVEBKUP(ℓ) does not probe ℓ irrespective of the outcomes of the probes, whereas OPT uses ℓ as backup in some paths of the decision tree and probes ℓ in other paths (Figure 1(a)). Owing to this difference, the probing sequence of OPT differs from that of RESERVEBKUP(ℓ) even in the paths where both use ℓ as backup (e.g., OPT and RESERVEBKUP(k) in Figure 1(a) and (b) respectively). Nevertheless, even in this example, for $\ell = k$, the gain of RESERVEBKUP(ℓ) is only 0.1% less than that of OPT.

4.3 Proof for Theorem 4.4

First observe that the optimal in the class of policies $\mathcal{P}(\ell)$ need not be unique. We consider OPT(ℓ) to be one optimal policy in $\mathcal{P}(\ell)$ that satisfies the following property. Suppose channel j is the last channel probed in a path in the decision tree of OPT(ℓ); let m be the node at which j is probed. Then, OPT(ℓ) would attain a lesser gain if it were not to probe j at m. Clearly, such optimal policies exist, and can be obtained by progressively modifying the decision tree of any optimal policy (that is by progressively removing the lowest node at which a channel is probed and which can be removed without reducing the gain of the tree).

We first observe the following about OPT(ℓ). Let the highest state of a probed channel be u when OPT(ℓ) terminates its probing process. Then OPT(ℓ) transmits in the probed channel if $r_u \geq \tilde{r}_\ell[0]$ and transmits in ℓ otherwise. Thus, the channel selection for RESERVEBKUP(ℓ) is optimal in $\mathcal{P}(\ell)$. We now state and prove three lemmas which will establish that the probing process for RESERVEBKUP(ℓ) is optimal in $\mathcal{P}(\ell)$, and thereby prove Theorem 4.4.

Lemma 4.6
1. *If all channels in $\bigcup_{v>\max(u,w_\ell-1)} H_{v,\ell}$ have already been probed, and the best state seen so far is u, then $OPT(\ell)$ does not probe any further.*
2. *$OPT(\ell)$ can not terminate the probing process when there an is un-probed channel in $\bigcup_{v>\max(u,w_\ell-1)} H_{v,\ell}$, and the best state seen so far is u.*

Proof: The first part of the lemma clearly holds for $u = K-1$ since the optimal policy does not probe any further after observing a channel in state $K-1$. Let the first part not hold for some $u < K-1$. Thus, although all channels in $\bigcup_{v>\max(u,w_\ell-1)} H_{v,\ell}$ have already been probed, OPT(ℓ) probes further channels. Let j be the last channel probed by OPT(ℓ) in one such path. Let j be probed at node m of the decision tree. Let all channels probed before probing j at m be in state q or lower where $q \geq u$. Note that after probing j OPT(ℓ) transmits in (a) backup ℓ if the maximum of q and j's state is $w_\ell - 1$ or lower and (b) the channel that was in state q otherwise. Now, consider another policy $C \in \mathcal{P}(\ell)$ which is similar to OPT(ℓ) except that it does not probe j at node m, and instead transmits in (a) backup ℓ if $q < w_\ell$ and (b) the probed channel that is in state q otherwise. Let Δ be the difference between the gains of OPT(ℓ) and C. We will arrive at a contradiction by showing that $\Delta \leq 0$. Hence, the first part of the lemma holds.

$$
\begin{aligned}
& \Delta \\
= & \sum_{k=0}^{K-1} p_{kj} \max(r_k, r_q, \tilde{r}_\ell[0]) - c_j - \max(r_q, \tilde{r}_\ell[0]) \\
= & \sum_{k=\max(q,w_\ell-1)+1}^{K-1} p_{kj} \left(\max(r_k, r_q, \tilde{r}_\ell[0]) - \max(\tilde{r}_\ell[0], r_q)\right) \\
& -c_j \\
= & \sum_{k=\max(q,w_\ell-1)+1}^{K-1} p_{kj} \left(r_k - \max(\tilde{r}_\ell[0], r_q)\right) - c_j.
\end{aligned}
$$

First, let $q \geq w_\ell - 1$. Thus, since $q \geq u$, $q+1 > \max(u, w_\ell - 1)$. Thus, $j \notin \cup_{v \geq q+1} H_{v,\ell}$. Thus, $\Delta < 0$. Now, let $q < w_\ell - 1$. Then, $r_q \leq r_{w_\ell - 1} \leq \tilde{r}_\ell[0]$. Thus, $\Delta = \sum_{k=w_\ell}^{K-1} p_{kj} \left(r_k - \max(\tilde{r}_\ell[0], r_{w_\ell-1})\right) - c_j$. Also, $u \leq q < w_\ell - 1$. Then, since $j \notin \bigcup_{v>\max(u,w_\ell-1)} H_{v,\ell}$, $j \notin \bigcup_{v \geq w_\ell} H_{v,\ell}$. Thus, $\Delta < 0$.

The second part of the lemma holds by vacuity when $u = K-1$ since $H_{K,\ell} = \phi$. Let $u < K-1$. Suppose $OPT(\ell)$ terminates the probing process even when there an is un-probed channel in $H_{v,\ell}$ for some $v > \max(u, w_\ell - 1)$, and the best state seen so far is u. Then, OPT(ℓ) transmits in (a) ℓ if $u < w_\ell$ and (b) a probed channel that is in state u otherwise. Consider another policy $D \in \mathcal{P}(\ell)$ which is similar to OPT(ℓ) except that it probes an additional channel $j \in H_{v,\ell}$, and transmits in (a) ℓ if the maximum of u and j's state is $w_\ell - 1$ or lower and (b) the probed channel that has the highest state otherwise. Let Δ_1 be the difference between the gains of OPT(ℓ) and D. We will show that $\Delta_1 < 0$ which is a contradiction. Hence, the second part of the lemma holds.

$$
\begin{aligned}
\Delta_1 &= \max(r_u, \tilde{r}_\ell[0]) - \sum_{k=0}^{K-1} p_{kj} \max(r_k, r_u, \tilde{r}_\ell[0]) + c_j \\
&= c_j - \sum_{k=\max(u,w_\ell-1)+1}^{K-1} p_{kj} \left(r_k - \max(\tilde{r}_\ell[0], r_u)\right) \\
&= c_j - \sum_{k=\max(u,w_\ell-1)+1}^{v-1} p_{kj} \left(r_k - \max(\tilde{r}_\ell[0], r_u)\right) \\
&\quad - \sum_{k=v}^{K-1} p_{kj} \left(r_k - \max(\tilde{r}_\ell[0], r_u)\right) \\
&\leq c_j - \sum_{k=v}^{K-1} p_{kj} \left(r_k - \max(\tilde{r}_\ell[0], r_{v-1})\right).
\end{aligned}
$$

The last inequality follows since $r_k \geq \max(\tilde{r}_\ell[0], r_u)$ for $k > \max(u, w_\ell - 1)$ and $r_{v-1} \geq r_u$ since $v \geq \max(u, w_\ell - 1) + 1$. The result follows since $j \in H_{v,\ell}$. $\square$

Lemma 4.7 *Let $w_\ell \leq u < K$.* OPT*(ℓ) probes all channels in $H_{u,\ell}$ before probing any channel that is not in $\cup_{v \geq u} H_{v,\ell}$. Also,* OPT*($\ell$) probes all channels of $H_{u,\ell}$ unless one of the probed channels is in state u or a higher state, and probes these channels in non-increasing order of $\tilde{r}_j[u] - \frac{c_j}{\tilde{p}_j[u]}$.*

Proof: Suppose the lemma does not hold. Then there exists a node in the decision tree of OPT(ℓ), which OPT(ℓ) visits with positive probability[3], such that the decisions at it violate the lemma. Let m be such a node which is also the farthest from the root node among those that violate this lemma. Then there exists a state $q \geq w_\ell$ such that there exists a channel in $H_{q,\ell}$ that has not been probed by OPT(ℓ) before it visits m and the best state seen so far is $q - 1$ or worse. Let u be the highest state that satisfies both the above criteria and let j be the channel with the largest $\tilde{r}_j[u] - \frac{c_j}{\tilde{p}_j[u]}$ value among the unprobed channels of $H_{u,\ell}$. At node m, OPT(ℓ) does not probe j contradicting the lemma. From the second part of lemma 4.6, the probing process of OPT(ℓ) can not terminate at m. Thus, OPT(ℓ) probes some channel $i \neq j$ at node m. Note that $i \notin \cup_{v>u} H_{v,\ell}$.

Since OPT(ℓ) have already probed all channels in $\cup_{v>u} H_{u,\ell}$, if channel i is in state u or a higher state, OPT(ℓ) does not probe any further (first part of lemma 4.6), and transmits in i (since i has the highest state, say s, among the probed channels and $r_s \geq r_u \geq r_{w_\ell} > \tilde{r}_\ell[0]$). If i is in a state $u - 1$ or a lower state, since $H_{u,\ell}$ has un-probed channels, the probing process can not terminate (second part of lemma 4.6). Now, OPT(ℓ) probes j next (otherwise m will not be the node farthest from the root to violate the lemma). Using similar arguments, it follows that after probing j, OPT(ℓ) transmits in j if j is in state u or a higher state.

The situation resembles the decision tree in Figure (2a). The trees $T_1 \ldots T_{u^2}$ correspond to observing the ordered pair $(i = u', j = u'')$ where $0 \leq u', u'' \leq u - 1$. The square boxes denote that OPT(ℓ) do not probe anything else.

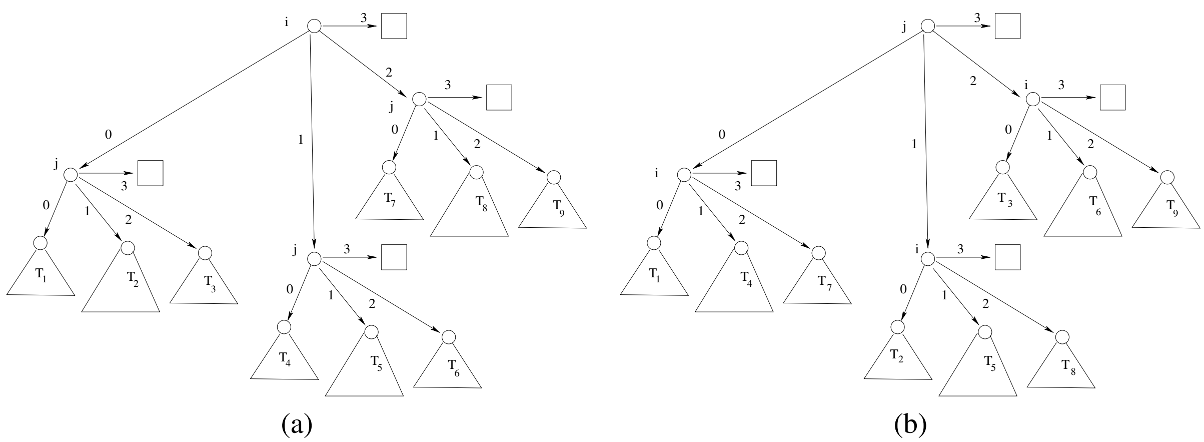


Figure 2: The decision trees of the Optimal policy for $u = 3$

Let OPT(ℓ) not traverse node m with probability α_1, traverse node m and stop after probing i or j with probability α_2, traverse node m and continue after probing j with probability α_3. By assumption, $\alpha_2 > 0$. Let the conditional expected gains given these scenarios be G_1, G_2, G_3 respectively. Then, the expected

[3]If the lemma is violated at a node which OPT(ℓ) visits with 0 probability, we can, without reducing the gain of OPT(ℓ), alter the decisions at the node so as to satisfy the lemma. Hence, without any loss of generality, we assume that the decisions of OPT(ℓ) satisfy the lemma at all such nodes.

gain of OPT(ℓ) is $G_{\text{OPT}(\ell)} = \sum_{i=1}^{3} \alpha_i G_i$. Let the total probing cost en-route to node m be C_1. Then, $G_2 = \frac{1-\alpha_1}{\alpha_2} \left(\tilde{p}_i[u]\tilde{r}_i[u] - c_i + (1-\tilde{p}_i[u])[\tilde{p}_j[u]\tilde{r}_j[u] - c_j] - C_1\right)$.

Now consider an alternate policy A that is similar to OPT(ℓ) except for the tree rooted at node m. Figure (2) shows the tree rooted at node m in policy A. Note that A probes j at node m and subsequently probes i unless j is in state u or a higher state. The tree T' corresponding to the ordered pair $(i = u', j = u'')$ is assigned appropriately, on the branch corresponding to observing j in u'' and subsequently observing i in u'. The contributions to the gain from the trees $T_1, \ldots T_{u^2}$ remain the same because in both the scenarios the probability of probing these trees are the same. Thus, the expected gain of this new policy is $G_C = \alpha_1 G_1 + \alpha_2 G_2' + \alpha_3 G_3$, where $G_2' = \frac{1-\alpha_1}{\alpha_2} \left(\tilde{p}_j[u]\tilde{r}_j[u] - c_j + (1-\tilde{p}_j[u])[\tilde{p}_i[u]\tilde{r}_i[u] - c_i] - C_1\right)$.

Now *if* $i \in H_{u,\ell}$ then we have $\tilde{r}_j[u] - \frac{c_j}{\tilde{p}_j[u]} > \tilde{r}_i[u] - \frac{c_i}{\tilde{p}_j[u]}$ which is the condition that arises from violating the non-increasing order. If $i \notin H_{u,\ell}$, then since $i \notin \cup_{v>u} H_{v,\ell}$ and $u \geq w_\ell$, $\tilde{r}_i[u] - \frac{c_i}{\tilde{p}_i[u]} \leq \max(r_{u-1}, \tilde{r}_\ell[0])$. But $\tilde{r}_j[u] - \frac{c_j}{\tilde{p}_j[u]} > \max(r_{u-1}, \tilde{r}_\ell[0])$ since $j \in H_{u,\ell}$ and $u \geq w_\ell$. Therefore, *in both cases* we have $\tilde{r}_j[u] - \frac{c_j}{\tilde{p}_j[u]} > \tilde{r}_i[u] - \frac{c_i}{\tilde{p}_i[u]}$. But this implies that

$$
\begin{aligned}
& \frac{G_C - G_{\text{OPT}(\ell)}}{1-\alpha_1} \\
= \; & \tilde{p}_j[u]\tilde{r}_j[u] - c_j + (1-\tilde{p}_j[u])\{\tilde{p}_i[u]\tilde{r}_i[u] - c_i\} \\
& -\tilde{p}_i[u]\tilde{r}_i[u] + c_i - (1-\tilde{p}_i[u])\{\tilde{p}_j[u]\tilde{r}_j[u] - c_j\} \\
= \; & \tilde{p}_i[u]\tilde{p}_j[u]\left(\tilde{r}_j[u] - \frac{c_j}{\tilde{p}_j[u]} - \tilde{r}_i[u] + \frac{c_i}{\tilde{p}_i[u]}\right) > 0.
\end{aligned}
$$

Thus, since $\alpha_1 < 1$, $G_C > G_{\text{OPT}(\ell)}$. Thus, we arrive at a contradiction. The result follows. □

Lemma 4.8 OPT*(ℓ) probes only channels in* $\cup_{v=w_\ell}^{K-1} H_{v,\ell}$.

Proof: From the first part of lemma 4.6, OPT(ℓ) terminates its probing process after probing all channels in $\cup_{v=w_\ell}^{K-1} H_{v,\ell}$. From lemma 4.7, OPT($\ell$) must probe all channels in $\cup_{v=w_\ell}^{K-1} H_{w_\ell,\ell}$ before probing a channel that is not in $\cup_{v=w_\ell}^{K-1} H_{v,\ell}$. The result follows. □

The optimality of the probing process for RESERVEBKUP(ℓ) in $\mathcal{P}(\ell)$ follows from lemmas 4.7 and 4.8. Thus, Theorem 4.4 follows.

5 The Unsaturated Sender Problem

In Subsection 5.1, we first present a stable policy UNSAT(ϵ) that attains a gain which is within a factor of $1-\epsilon$ of the optimum gain. The computation time of UNSAT(ϵ) is however exponential in n for all ϵ. We next present another stable policy UNSATAPPROX(ϵ) that approximates the optimal within a factor of $2/3$ for small ϵ; the computation time for UNSATAPPROX(ϵ) is polynomial in the input size. In Subsection 5.3, we prove the performance guarantee of UNSATAPPROX(ϵ).

5.1 Approximately Optimal Unsaturated Policies

We now characterize UNSAT(ϵ) that is stable and attains a gain which is within a factor of $1-\epsilon$ of the optimum gain.

Since the number of channels and the number of transmission states of the channels are both finite, number of possible probing, selection and transmission decisions are also finite in any given slot. The decision trees representing these decisions constitute a finite set Π. For example, the decision trees in Figure 1(a),(b) are examples of decision trees in Π. In addition, if the decision trees in Figure 1(b) are modified so as not to transmit at the end of the left-most paths, the modifications will also constitute decision trees in Π. Note that Π also includes the decision tree that neither probes nor transmits in any channel.

Definition 5.1 *Let C_σ denote the expected probing cost in decision tree $\sigma \in \Pi$, and $\mathcal{M}_\sigma$ be the set of leaf nodes where the decision is to transmit. Let $\hat{p}_{m\sigma}$ denote the probability that the leaf node m is reached in σ, and if $m \in \mathcal{M}_\sigma$, $\hat{r}_{m\sigma}$ is the probability of success for the transmission at m. Let $\mathcal{S}_\sigma = \sum_{m\in\mathcal{M}_\sigma} \hat{p}_{m\sigma}$ and $\mathcal{G}_\sigma = \sum_{m\in\mathcal{M}_\sigma} \hat{r}_{m\sigma}\hat{p}_{m\sigma} - C_\sigma$ for a $\sigma \in \Pi$. Let $\delta(\epsilon) = \min(\epsilon, (1-\lambda)/2)$.*

Let $\{\beta^*(\epsilon)\}$ be the probability distribution obtained as a solution of the following linear program, denoted as LPUNSAT(ϵ).

$$\text{Maximize} \sum_{\sigma\in\Pi} \beta_\sigma \mathcal{G}_\sigma$$

$$\begin{aligned} \textstyle\sum_{\sigma\in\Pi} \beta_\sigma \mathcal{S}_\sigma &= \lambda + \delta(\epsilon) \qquad (\textit{stability constraint}) \\ \textstyle\sum_{\sigma\in\Pi} \beta_\sigma &= 1 \\ \beta_\sigma &\geq 0 \;\forall\; \sigma \in \Pi \end{aligned}$$

Policy UNSAT(ϵ)

In each slot, select a $\sigma \in \Pi$ in accordance with the probability distribution $\{\beta^*(\epsilon)\}$.
If the slot is busy, probe channels, decide whether to transmit, and select a channel as per σ.

Let $Q^*(\epsilon)$ denote the optimal value of the objective function of LPUNSAT(ϵ).

Theorem 5.1 *For any $\epsilon > 0$ and $\lambda \in [0, 1)$, UNSAT($\epsilon\lambda$) is stable and attains a gain of at least $\frac{G_{UnSat}}{1+\epsilon}$. Furthermore, $G_{UnSat} \leq Q^*(\epsilon)$.*

Proof: In any busy slot, UNSATAPPROX($\epsilon\lambda$) transmits a packet with probability $\beta^*(\epsilon\lambda)$. Note that $\beta^*(\epsilon\lambda)$ is chosen so that this probability equals $\lambda + \delta(\epsilon\lambda)$ which exceeds λ. Since UNSATAPPROX($\epsilon\lambda$) constitutes a Markov chain, stability follows from standard results and analytical techniques (Theorem 2.2.3 in [10], [7]).

Next consider the expected gain of the policy LPUNSAT($\epsilon\lambda$). Since the system is stable and LPUNSAT($\epsilon\lambda$) transmits a packet with probability $\lambda + \delta(\epsilon\lambda)$ in each busy slot and $\delta(\epsilon\lambda) \leq \epsilon\lambda$, using Little's law, at least $\frac{1}{1+\epsilon}$ of slots are busy. The gain of LPUNSAT($\epsilon\lambda$) in each busy slot is $Q^*(\epsilon\lambda)$. Thus, LPUNSAT($\epsilon\lambda$) attains a gain of at least $Q^*(\epsilon\lambda)/(1+\epsilon)$.

We will finally show that $G_{\text{UnSat}} \leq Q^*(\epsilon\lambda)$, completing the proof. We will assume the optimal policy UNSAT is ergodic. Let the steady state rate at which it chooses policy σ for execution be denoted β_σ. Since the policy is stable, we must have: $\sum_{\sigma\in\Pi} \beta_\sigma \mathcal{S}_\sigma = \lambda$. The expected gain of this policy is simply $\sum_{\sigma\in\Pi} \beta_\sigma \mathcal{G}_\sigma$. The steady state rates therefore form a feasible solution to LPUNSAT(0). Therefore, $G_{\text{UnSat}} \leq Q^*(0)$. It is easy to see that $Q^*(0) \leq Q^*(\epsilon\lambda)$. This completes the proof of the theorem. □

The above theorem in retrospect: It would be useful to summarize the intuition behind the above result. If in a busy slot the sender takes actions as per the decision tree σ, it transmits with probability $\mathcal{S}_\sigma$ in the slot and it attains a gain of $\mathcal{G}_\sigma$ in the slot. Since the sender selects σ with a probability $\beta^*(\epsilon)_\sigma$ in any slot, the stability constraint in LPUNSAT(ϵ) therefore ensures that UNSAT(ϵ) offers a service rate of $\lambda + \delta(\epsilon)$ where $\delta(\epsilon) > 0$. Thus, the sender's queue is stable. Subject to this constraint, UNSAT(ϵ) chooses $\{\beta^*(\epsilon)\}$ so as to maximize its gain; this maximum turns out to be within a $1-\epsilon$ factor of the maximum gain.

Note that $|\Pi| \geq K^n$. Thus, LPUNSAT(ϵ) has an exponential number of variables, and the computation time for UNSAT(ϵ) is exponential in n for any ϵ. We next present a computationally simple policy, UNSATAPPROX(ϵ) that (a) attains approximately $2/3$ the gain of the optimal policy for small ϵ, (b) stabilizes the system for all positive ϵ and (c) can be computed in $O(n^2 K\tau)$ time, where τ is polynomial in the input size.

5.2 A Polynomial Time Approximation Algorithm

We first describe the following collection of threshold type policies RESERVEBKUP(ℓ, x) which will be used to construct UNSATAPPROX(ϵ). We first generalize the definition for $H_{u,\ell}$ as follows.

Definition 5.2 *Let $H_{u,\ell,x} = \phi$ for all $u \geq K$. For each ℓ, starting from $u = K-1$, down to $u = w_\ell$, recursively, define $H_{u,\ell,x} = \left\{ i | i \notin \bigcup_{v:v>u} H_{v,\ell,x},\ \text{and}\ \tilde{r}_i[u] - \frac{c_i}{\tilde{p}_i[u]} > \max(\tilde{r}_\ell[0], r_{u-1}, x) \right\} \setminus \{\ell\}$. Assume $c_i/\tilde{p}_i[u] = +\infty$ when $\tilde{p}_i[u] = 0$.*

> RESERVEBKUP(ℓ, x)
>
> 1. Let $w_{\ell,x} = \min_u\{u : u \geq 0, r_u > \max(\tilde{r}_\ell[0], x)\}$. If $r_u \leq \max(\tilde{r}_\ell[0], x)$ for all u, $w_{\ell,x} = K$.
> 2. Consider each $H_{u,\ell,x}$ in decreasing order of u starting from $u = K$ down to $u = w_{\ell,x}$.
> 3. (**Probing Process:**) Within each $H_{u,\ell,x}$ probe in non-increasing order of $\tilde{r}_j[u] - \frac{c_j}{\tilde{p}_j[u]}$, and stop if any channel is found to be in state u or above.
> 4. (**Selection (and decision to transmit):**) Let channel j have the highest state, say y, among all probed channels. If no channel have been probed $y = -1$. If $\max(r_y, \tilde{r}_\ell[0]) < x$, do not transmit. If $\max(r_y, \tilde{r}_\ell[0]) \geq x$, transmit in j if $r_y \geq \tilde{r}_\ell[0]$ and transmit in ℓ otherwise.

Note that RESERVEBKUP(ℓ, x) is similar to RESERVEBKUP(ℓ) except that RESERVEBKUP(ℓ, x) selects a transmission threshold, x, apriori, and transmits only if a probed channel is in state j or a higher state such that $r_j \geq x$ or if the probability of success of the backup channel ℓ is not lower than x. Therefore, RESERVEBKUP(ℓ, x), probes only those channels for which the expected rewards conditioned on being in states k and above, where $r_k > x$, exceed the probing cost.

Let σ_x denote the decision tree for RESERVEBKUP(ℓ, x) for that ℓ for which it attains the maximum gain among all $\ell \in \{0, \ldots, n\}$. UNSATAPPROX($\epsilon$) computes two transmission thresholds $x^+(\epsilon), x^-(\epsilon)$ and a probability $\hat{\beta}(\epsilon)$. Subsequently, in each slot it selects $\sigma_{x^+(\epsilon)}$ and $\sigma_{x^-(\epsilon)}$ with probabilities $\hat{\beta}(\epsilon)$ and $1-\hat{\beta}(\epsilon)$ respectively, and acts in accordance with the selected decision tree. The parameters $x^+(\epsilon), x^-(\epsilon), \hat{\beta}(\epsilon)$ are selected so as to ensure that UNSATAPPROX(ϵ) is stable and attains at least $\frac{2(1-\epsilon)}{3(1+\epsilon)}$ times the gain of the optimal policy. Formal description of UNSATAPPROX(ϵ) follows.

> UNSATAPPROX(ϵ)
>
> Determine $x^+(\epsilon), x^-(\epsilon), \hat{\beta}(\epsilon)$ using SEARCH(ϵ) (to be described later). Compute $\sigma_{x^+(\epsilon)}$ and $\sigma_{x^-(\epsilon)}$ using procedures RESERVEBKUP(ℓ, x) for $\ell \in \{0, \ldots, n\}$ and $x \in \{x^+(\epsilon), x^-(\epsilon)\}$.
> In every slot the sender selects $\sigma_{x^+(\epsilon)}$ and $\sigma_{x^-(\epsilon)}$ with probabilities $\hat{\beta}(\epsilon)$ and $1-\hat{\beta}(\epsilon)$ respectively.
> If the sender's queue is non-empty, it decides its actions in accordance with the selected σ.

We now describe how the parameters $x^+(\epsilon), x^-(\epsilon)$ and $\hat{\beta}(\epsilon)$ are selected. Let $\ell^* = \arg\max_\ell \tilde{r}_\ell[0]$. In each busy slot in which the sender acts in accordance with the decision tree $\sigma \in \Pi$, it transmits a packet with probability $\mathcal{S}_\sigma$. Now, $\mathcal{S}_{\sigma_x} = 1$ for $x = 0$ and $\mathcal{S}_{\sigma_x} = 0$ for $x = 2r_{K-1}$. Furthermore, $\mathcal{S}_{\sigma_x}$ does not increase with increase of x. Thus, there exists $x^+(\epsilon), x^-(\epsilon) \in [0, 2r_{K-1}]$ such that $0 \leq x^+(\epsilon) - x^-(\epsilon) \leq (2/3)\epsilon\tilde{r}_{\ell^*}[0]$ and $\mathcal{S}_{\sigma_{x^+(\epsilon)}} \leq \lambda + \delta(\epsilon) \leq \mathcal{S}_{\sigma_{x^-(\epsilon)}}$. UNSATAPPROX($\epsilon$) determines such a pair, and also the probability $\hat{\beta}(\epsilon)$, using a binary search strategy which we describe next.

SEARCH(ϵ)

Let $x^+(\epsilon) = 2r_{K-1}, x^-(\epsilon) = 0$.
Let temp $= (x^+(\epsilon) + x^-(\epsilon))/2$ (*Step 1*).
If $\left(\mathcal{S}_{\text{temp}} = \lambda + \delta(\epsilon)\right)$, $x^+(\epsilon) = x^-(\epsilon) =$ temp, $\hat{\beta}(\epsilon) = 1$, and terminate the search.
If $\left(\mathcal{S}_{\text{temp}} < \lambda + \delta(\epsilon)\right)$, $x^+(\epsilon) =$ temp. If $\left(\mathcal{S}_{\text{temp}} > \lambda + \delta(\epsilon)\right)$, $x^-(\epsilon) =$ temp.
If $x^+(\epsilon) - x^-(\epsilon) \leq 2\epsilon\tilde{r}_{\ell^*}[0]/3$, $\hat{\beta}(\epsilon) = \frac{\mathcal{S}_{\sigma_{x^-(\epsilon)}} - \lambda - \delta(\epsilon)}{\mathcal{S}_{\sigma_{x^-(\epsilon)}} - \mathcal{S}_{\sigma_{x^+(\epsilon)}}}$, and terminate the search. Otherwise, return to *Step 1*.

We first prove the stability of UNSATAPPROX(ϵ), and subsequently focus on the performance guarantee.

Proposition 5.1 *For any $\epsilon > 0$, $\lambda \in [0, 1)$, UNSATAPPROX(ϵ) is stable.*

Proof: In any busy slot, UNSATAPPROX(ϵ) transmits a packet with probability $\hat{\beta}\mathcal{S}_{\sigma_{x^+(\epsilon)}} + (1 - \hat{\beta})\mathcal{S}_{\sigma_{x^-(\epsilon)}}$. Note that $\hat{\beta}$ is chosen so that this probability equals $\lambda + \delta(\epsilon)$ which exceeds λ. Since UNSATAPPROX(ϵ) constitutes a Markov chain, stability follows from standard results and analytical techniques (Theorem 2.2.3 in [10], [7]). □

Theorem 5.2 *For any $\epsilon > 0$, and $\lambda \in [0, 1)$, UNSATAPPROX($\epsilon\lambda$) (already shown to be stable) attains a gain which is at least $\frac{2(1-\epsilon)G_{UnSat}}{3(1+\epsilon)}$.*

We prove Theorem 5.2 in the next subsection. We now determine the computation time for UNSATAPPROX(ϵ). Note that the number of iterations in SEARCH(ϵ) is at most $\log(3r_{K-1}/\epsilon\tilde{r}_{\ell^*}[0])$. In each iteration, SEARCH(ϵ) determines $\sigma_{x^+(\epsilon)}$ and $\sigma_{x^-(\epsilon)}$. This requires computations of the gains of RESERVEBKUP(ℓ, x) for $\ell \in \{0, \ldots, n\}$ and $x \in \{x^+(\epsilon), x^-(\epsilon)\}$, which require $O(n^2K)$ time. Subsequently, $\mathcal{S}_{\sigma^x}$ can be computed in $O(n)$ time for $x \in \{x^+(\epsilon), x^-(\epsilon)\}$. Thus, the overall computation time is $O\left(n^2K\log(3r_{K-1}/\epsilon\lambda\tilde{r}_{\ell^*}[0])\right)$ which is polynomial in the input size. Finally, UNSATAPPROX($\epsilon\lambda$) can be stored using $O(n)$ space and executed in $O(n)$ time.

5.3 Proof for Theorem 5.2

The proof for Theorem 5.2 depends on lemmas 5.3 to 5.6; we state and prove these first. The key is to compute a lower bound of the optimal strategy, and we revisit the linear program LPUNSAT(ϵ). We now consider the linear program LPRELAXATION(ϵ) which is the Lagrangian relaxation of LPUNSAT(ϵ). The quantities $\beta_\sigma, \mathcal{G}_\sigma, \delta(\epsilon)$ are the same as in Definition 5.1.

$$\text{Maximize} \ \sum_\sigma \beta_\sigma \mathcal{G}_\sigma - \mathcal{L}\left(\sum_\sigma \beta_\sigma \mathcal{S}_\sigma - \lambda - \delta(\epsilon)\right)$$

$$\begin{aligned} \textstyle\sum_{\sigma \in \Pi} \beta_\sigma &= 1 \\ \beta_\sigma &\geq 0 \ \forall \ \sigma \in \Pi \end{aligned}$$

The parameter $\mathcal{L}$ is the Lagrange multiplier. Note that for any $\mathcal{L}$ the solution of LPRELAXATION(ϵ) is a *single decision tree*, $\sigma^*_{\mathcal{L}}$ that maximizes $\mathcal{G}_\sigma - \mathcal{L}\mathcal{S}_\sigma$ among all $\sigma \in \Pi$.

Lemma 5.3 *Let there exist a $\mathcal{L}$, $0 < c \leq 1$, and $\hat{\sigma}_{\mathcal{L}} \in \Pi$ such that*

$$\begin{aligned} \mathcal{G}_{\hat{\sigma}_{\mathcal{L}}} - \mathcal{L}\mathcal{S}_{\hat{\sigma}_{\mathcal{L}}} &\geq c\left(\mathcal{G}_{\sigma^*_{\mathcal{L}}} - \mathcal{L}\mathcal{S}_{\sigma^*_{\mathcal{L}}}\right), & (5)\\ \textit{and } \mathcal{S}_{\hat{\sigma}_{\mathcal{L}}} &= \lambda + \delta(\epsilon). & (6) \end{aligned}$$

Then $\mathcal{G}_{\hat{\sigma}_{\mathcal{L}}} \geq cQ^*(\epsilon)$.

In view of lemma 5.3, if the sender decides its actions in accordance with $\hat{\sigma}_{\mathcal{L}} \in \Pi$ in every busy slot, it attains a gain of at least $cQ^*(\epsilon)$ in every such slot. Thus, if a large fraction of slots are busy, its overall gain will be very close to $cQ^*(\epsilon)$. Since the maximum gain is upper bounded by $Q^*(\epsilon)$ (Theorem 5.1), the sender's gain is at least c times that of the optimum. We now prove the lemma.

Proof:

Consider an optimum solution $\{\beta^*(\epsilon)\}$ for LPUNSAT(ϵ).

$$\begin{aligned} Q^*(\epsilon) &= \sum_{\sigma\in\Pi} \beta^*(\epsilon)_\sigma \mathcal{G}_\sigma \\ &= \sum_{\sigma\in\Pi} \beta^*(\epsilon)_\sigma \mathcal{G}_\sigma - \mathcal{L}(\sum_{\sigma\in\Pi} \beta^*(\epsilon)_\sigma \mathcal{S}_\sigma - \lambda - \delta(\epsilon)) \\ &\quad \text{(from the stability constraint of LPUNSAT}(\epsilon)\text{).} \\ &= \sum_{\sigma\in\Pi} \beta^*(\epsilon)_\sigma \left(\mathcal{G}_\sigma - \mathcal{L}\mathcal{S}_\sigma\right) + \mathcal{L}(\lambda + \delta(\epsilon)) \\ &\leq \mathcal{G}_{\sigma^*_{\mathcal{L}}} - \mathcal{L}\mathcal{S}_{\sigma^*_{\mathcal{L}}} + \mathcal{L}(\lambda + \delta(\epsilon)) \\ &\leq \frac{1}{c}\left(\mathcal{G}_{\hat{\sigma}_{\mathcal{L}}} - \mathcal{L}\mathcal{S}_{\hat{\sigma}_{\mathcal{L}}}\right) + \mathcal{L}(\lambda + \delta(\epsilon)) \ \text{(from (5))} \qquad (7)\\ &\leq \mathcal{G}_{\hat{\sigma}_{\mathcal{L}}}/c - \mathcal{L}\left(\mathcal{S}_{\hat{\sigma}_{\mathcal{L}}} - \lambda - \delta(\epsilon)\right) \ \text{(since } c \in (0,1]) \\ &= \mathcal{G}_{\hat{\sigma}_{\mathcal{L}}}/c \ \text{(from (6)).} \end{aligned}$$

□

Given any $c \in (0,1]$, $\lambda \in [0,1)$, and $\epsilon > 0$ it is not clear (a) whether there exists a $\mathcal{L}$ and $\hat{\sigma}_{\mathcal{L}} \in \Pi$ that satisfy conditions (5) and (6) of Lemma 5.3 and (b) how to find such $\mathcal{L}, \hat{\sigma}_{\mathcal{L}} \in \Pi$. We now state and prove a similar lemma, lemma 5.4, which requires a different set of conditions for the desired $\hat{\sigma}_{\mathcal{L}} \in \Pi$. These conditions are always satisfied by some $\sigma \in \Pi$ (for $c = 2/3$) and these σ s can be easily computed. Lemmas 5.3, 5.4 will therefore lead to the $2/3$ approximation guarantee in Theorem 5.2.

Lemma 5.4 *Let $\lambda \in [0,1)$, $\epsilon > 0$. Let there exist $\mathcal{L}^+, \mathcal{L}^-$, $0 < c \leq 1$, and $\hat{\sigma}_{\mathcal{L}^+}, \hat{\sigma}_{\mathcal{L}^-} \in \Pi$ such that $\forall\, \mathcal{L} \in \{\mathcal{L}^+, \mathcal{L}^-\}$,*

$$\mathcal{G}_{\hat{\sigma}_{\mathcal{L}}} - \mathcal{L}\mathcal{S}_{\hat{\sigma}_{\mathcal{L}}} \geq c\left(\mathcal{G}_{\sigma^*_{\mathcal{L}}} - \mathcal{L}\mathcal{S}_{\sigma^*_{\mathcal{L}}}\right). \qquad (8)$$

$$\textit{Also, } \mathcal{S}_{\hat{\sigma}_{\mathcal{L}^+}} < \lambda + \delta(\epsilon) < \mathcal{S}_{\hat{\sigma}_{\mathcal{L}^-}}. \qquad (9)$$

$$\textit{Finally, } \mathcal{L}^- \geq 0,\ 0 < \mathcal{L}^+ - \mathcal{L}^- \leq c\epsilon Q^*(\epsilon)/(\lambda + \delta(\epsilon)). \qquad (10)$$

Consider $\{\beta\}$ such that $\beta_\sigma = \alpha$ for $\sigma = \hat{\sigma}_{\mathcal{L}^+}$, $\beta_\sigma = 1-\alpha$ for $\sigma = \hat{\sigma}_{\mathcal{L}^+}$, and $\beta_\sigma = 0$ for $\sigma \in \Pi\backslash\{\hat{\sigma}_{\mathcal{L}^+}, \hat{\sigma}_{\mathcal{L}^-}\}$.

1. *If $\alpha = \frac{\mathcal{S}_{\hat{\sigma}_{\mathcal{L}^-}} - \lambda - \delta(\epsilon)}{\mathcal{S}_{\hat{\sigma}_{\mathcal{L}^-}} - \mathcal{S}_{\hat{\sigma}_{\mathcal{L}^+}}}$, $\{\beta\}$ constitutes a feasible solution for* LPUNSAT(ϵ).
2. *$\sum_{\sigma\in\{\hat{\sigma}_{\mathcal{L}^+}, \hat{\sigma}_{\mathcal{L}^-}\}} \beta_\sigma \mathcal{G}_\sigma \geq c(1-\epsilon)Q^*(\epsilon)$.*

Now, if in every slot, the sender selects either $\hat{\sigma}_{\mathcal{L}^+}$ or $\hat{\sigma}_{\mathcal{L}^-}$ (the former with probability $\beta_{\hat{\sigma}_{\mathcal{L}^+}}$), and acts in accordance with the selected σ if it has a packet to transmit, it attains a gain of at least $c(1-\epsilon)Q^*(\epsilon)$. Thus, the maximum gain can be approximated within a factor close to c. The proof follows.

Proof: Using (9), it can be easily verified that $\{\beta\}$ satisfies all the constraints in LPUNSAT(ϵ). The second part of the lemma follows if we show that $Q^*(\epsilon) \leq \frac{1}{c(1-\epsilon)}\left(\beta_{\hat{\sigma}_{\mathcal{L}^-}}\mathcal{S}_{\hat{\sigma}_{\mathcal{L}^-}} + \beta_{\hat{\sigma}_{\mathcal{L}^+}}\mathcal{S}_{\hat{\sigma}_{\mathcal{L}^+}}\right)$.

Similar to the deduction of (7) and using (8) and the facts that $c \in (0,1]$ and $\mathcal{L}^+ > \mathcal{L}^- \geq 0$, we can show that

$$Q^*(\epsilon) \leq \frac{\mathcal{G}_{\hat{\sigma}_{\mathcal{L}^+}} - \mathcal{L}^+(\mathcal{S}_{\hat{\sigma}_{\mathcal{L}^+}} - \lambda - \delta(\epsilon))}{c}, \tag{11}$$

$$Q^*(\epsilon) \leq \frac{\mathcal{G}_{\hat{\sigma}_{\mathcal{L}^-}} - \mathcal{L}^-(\mathcal{S}_{\hat{\sigma}_{\mathcal{L}^-}} - \lambda - \delta(\epsilon))}{c}. \tag{12}$$

Note that $\beta_{\hat{\sigma}_{\mathcal{L}^+}} + \beta_{\hat{\sigma}_{\mathcal{L}^-}} = 1$. Thus, multiplying (11) with $\beta_{\hat{\sigma}_{\mathcal{L}^+}}$ and (12) with $\beta_{\hat{\sigma}_{\mathcal{L}^-}}$ and adding the products we get

$$\begin{aligned}
Q^*(\epsilon) &\leq \frac{1}{c}\left(\beta_{\hat{\sigma}_{\mathcal{L}^+}}\mathcal{G}_{\hat{\sigma}_{\mathcal{L}^+}} - \beta_{\hat{\sigma}_{\mathcal{L}^+}}\mathcal{L}^+(\mathcal{S}_{\hat{\sigma}_{\mathcal{L}^+}} - \lambda - \delta(\epsilon))\right) \\
&\quad + \frac{1}{c}\left(\beta_{\hat{\sigma}_{\mathcal{L}^-}}\mathcal{G}_{\hat{\sigma}_{\mathcal{L}^-}} - \beta_{\hat{\sigma}_{\mathcal{L}^-}}\mathcal{L}^-(\mathcal{S}_{\hat{\sigma}_{\mathcal{L}^-}} - \lambda - \delta(\epsilon))\right) \\
&= \frac{1}{c}\left(\beta_{\hat{\sigma}_{\mathcal{L}^+}}\mathcal{G}_{\hat{\sigma}_{\mathcal{L}^+}} + \beta_{\hat{\sigma}_{\mathcal{L}^-}}\mathcal{G}_{\hat{\sigma}_{\mathcal{L}^-}}\right) \\
&\quad + \frac{1}{c}\beta_{\hat{\sigma}_{\mathcal{L}^+}}(\mathcal{L}^+ - \mathcal{L}^-)(\lambda + \delta(\epsilon) - \mathcal{S}_{\hat{\sigma}_{\mathcal{L}^+}}) \\
&\quad - \frac{1}{c}\mathcal{L}^-(\beta_{\hat{\sigma}_{\mathcal{L}^+}}\mathcal{S}_{\hat{\sigma}_{\mathcal{L}^+}} + \beta_{\hat{\sigma}_{\mathcal{L}^-}}\mathcal{S}_{\hat{\sigma}_{\mathcal{L}^-}} - \lambda - \delta(\epsilon)) \\
&= \frac{1}{c}\left(\beta_{\hat{\sigma}_{\mathcal{L}^+}}\mathcal{G}_{\hat{\sigma}_{\mathcal{L}^+}} + \beta_{\hat{\sigma}_{\mathcal{L}^-}}\mathcal{G}_{\hat{\sigma}_{\mathcal{L}^-}}\right) \\
&\quad + \frac{1}{c}\beta_{\hat{\sigma}_{\mathcal{L}^+}}(\mathcal{L}^+ - \mathcal{L}^-)(\lambda + \delta(\epsilon) - \mathcal{S}_{\hat{\sigma}_{\mathcal{L}^+}}) \\
&\quad \text{(from the stability constraint for } \{\beta\}\text{)} \\
&\leq \frac{1}{c}\left(\beta_{\hat{\sigma}_{\mathcal{L}^+}}\mathcal{G}_{\hat{\sigma}_{\mathcal{L}^+}} + \beta_{\hat{\sigma}_{\mathcal{L}^-}}\mathcal{G}_{\hat{\sigma}_{\mathcal{L}^-}}\right) \\
&\quad + \frac{1}{c}(\mathcal{L}^+ - \mathcal{L}^-)(\lambda + \delta(\epsilon))
\end{aligned}$$

Now, the result follows from (10). □

Recall that σ_x is the decision tree for RESERVEBKUP(ℓ, x) that attains the maximum gain among all $\ell \in \{0, \ldots, n\}$. We now prove that for any $\lambda \in [0,1), \epsilon > 0$, $c = 2/3$, and for any x, σ_x satisfies the conditions in (5), (8) in lemmas 5.3 and 5.4 respectively. This explains why UNSATAPPROX(ϵ) in each slot selects σ_x for some x. Before we prove the gain we achieve; we would need to prove a structure theorem similar to the Structure Theorem 4.2.

Theorem 5.5 (Generalized Structure Theorem) *Consider a fictitious system where the sender is saturated and receives a reward of $r_j - x$ whenever it transmits in a channel which is in state j. There exists an optimum policy in the fictitious system that uses a unique backup channel whenever it transmits in a backup channel. The backup channel is used at the end of one $\leq i$ path.*

Proof: Observe that Lemma 4.1 holds in this fictitious system. The proof is the same as that in the original system.

Let an optimum policy OPT3 use backup channels at the end of at least two distinct paths P_1 and P_2 in the decision tree. Let P_1 and P_2 intersect at node m in the decision tree. Then, OPT3 probes a channel j at m and uses P_1 and P_2 for two different states of j. Let k be the higher of the two states. Then, by lemma 4.1, there exists another optimum policy OPT4 which differs from OPT3 only in the decision tree rooted at m which is a $\leq k$ tree in OPT4. Clearly, OPT4 has strictly fewer number of paths ending in backups. The Theorem follows from recursive usage of this argument. □

Lemma 5.6 *For any* $\lambda \in [0, 1), \epsilon > 0$, *and for any real* x, $\mathcal{G}_{\sigma_x} - x\mathcal{S}_{\sigma_x} \geq 2/3 \left(\mathcal{G}_{\sigma_x^*} - x\mathcal{S}_{\sigma_x^*}\right)$.

Proof: Consider again the fictitious system where the sender is saturated and receives a reward of $r_j - x$ whenever it transmits in a channel which is in state j. Thus, transmission in a backup channel ℓ in this system fetches a gain of $\tilde{r}_\ell[0] - x$. The set of decision trees in this system is Π, irrespective of x. Note that the gain of any $\sigma \in \Pi$ in this system is $\mathcal{G}_\sigma - x\mathcal{S}_\sigma$ and depends on x. Since by definition σ_x^* maximizes $\mathcal{G}_\sigma - x\mathcal{S}_\sigma$ among all $\sigma \in \Pi$, σ_x^* attains the maximum gain in this system. Let $F = \mathcal{G}_{\sigma_x} - x\mathcal{S}_{\sigma_x}$ and $BEST = \mathcal{G}_{\sigma_x^*} - x\mathcal{S}_{\sigma_x^*}$.

In this fictitious system, for each x, multiple σ may maximize the gain. The generalized structure theorem (Theorem 5.5) shows that for each x at least one σ_x^* uses a unique backup channel whenever a backup channel is used for transmission. We therefore consider a σ_x^* that uses a unique backup, say ℓ. Let σ_x^* use ℓ as backup with probability α. Let $R = \sum_{i:r_i \geq x} p_{i\ell}(r_i - x)$ and $T = \sum_{i:r_i < x} p_{i\ell}(x - r_i)$.

Let $\mathcal{P}'(j)$ be the set of all decision trees in Π that use no channel other than j as a backup. Note that policies in $\mathcal{P}'(0)$ never transmit in any backup channel. Using similar arguments as in the proof of Theorem 4.4, it can be shown that in this fictitious system RESERVEBKUP(j, x) attains the maximum gain among all policies in $\mathcal{P}'(j)$. Thus, F is the maximum gain attained in this fictitious system by any policy in $\cup_{j=0}^{n} \mathcal{P}'(j)$.

Now, construct a policy σ_1 that is similar to σ_x^* except that whenever σ_x^* transmits in ℓ, σ_1 does not transmit. Thus σ_1 attains a gain of at least BEST - $\alpha(R - T)$. Also, $\sigma_1 \in \mathcal{P}'(0)$. Thus, its gain is upper bounded by F. Thus $F \geq BEST - \alpha(R - T)$. Now, consider the policy that transmits in ℓ every slot without probing. This policy is in $\mathcal{P}'(j)$ and attains a gain of $R - T$. Thus, $F \geq R - T$. It follows that

$$(1 + \alpha)F \geq BEST. \tag{13}$$

Next, construct another policy σ_2 that is similar to σ_x^* except that σ_2 never probes ℓ; instead wherever σ_x^* probes ℓ, σ_2 follows the same course of actions as σ_x^* does after discovering ℓ in state 0. The gain of σ_2 is at least $BEST - (1 - \alpha)(R - c_\ell)$, because if ℓ was in state i such that $r_i < x$, σ_x^* will never transmit in j. Also, $\sigma_2 \in \mathcal{P}'(\ell)$. Therefore $F \geq BEST - (1 - \alpha)(R - c_\ell)$. Now consider another policy σ_3 that probes ℓ and subsequently transmits only if ℓ is in a state i such that $r_i \geq x$; σ_3 neither probes nor transmits in any other channel. Clearly, σ_3 attains a gain of $R - c_\ell$. Also, $\sigma_3 \in \mathcal{P}'(0)$. Thus, $F \geq R - c_\ell$. Combining the last two equations,

$$(2 - \alpha)F \geq BEST. \tag{14}$$

Adding (13) and (14), we get $F \geq 2/3BEST$.

□

Proof: **(Of Theorem 5.2.)** Let $\epsilon_1 = \lambda\epsilon$. We have argued that for all $\lambda \in [0, 1)$ and $\epsilon_1 > 0$, UNSATAPPROX(ϵ_1) terminates in a finite number of iterations (last paragraph of Section 5.1).

Clearly, UNSATAPPROX(ϵ_1) transmits a packet with probability $\hat{\beta}(\epsilon_1)\mathcal{S}_{\sigma_{x^+(\epsilon_1)}} + (1 - \hat{\beta}(\epsilon_1))\mathcal{S}_{\sigma_{x^-(\epsilon_1)}}$ in each busy slot, and by the choice of $\hat{\beta}(\epsilon_1)$ this probability equals $\lambda + \delta(\epsilon_1)$. Since the system is stable, using Little's law, $\frac{\lambda}{\lambda+\epsilon_1}$ of slots are busy. The gain of UNSATAPPROX(ϵ_1) in each busy slot is $A = \hat{\beta}(\epsilon_1)\mathcal{G}_{\sigma_{x^+(\epsilon_1)}} +$

$(1-\hat{\beta}(\epsilon_1))\mathcal{G}_{\sigma_{x^-(\epsilon_1)}}$. Thus, $G_{\text{UnSatApprox}(\epsilon_1)} = \frac{A}{1+\epsilon}$. Since $Q^*(\epsilon_1) \geq G_{\text{UnSat}}$ (Theorem 5.1), the result follows if we can show that $A \geq 2(1-\epsilon)Q^*(\epsilon_1)/3$.

Let $x^+(\epsilon_1) = x^-(\epsilon_1) = x$. Then $A = \mathcal{G}_{\sigma_x}$ and $\mathcal{S}_{\sigma_x} = \lambda + \delta(\epsilon_1)$. From lemma 5.6, (5) and (6) of lemma 5.3 hold with $\mathcal{L} = x, \hat{\sigma_{\mathcal{L}}} = \sigma_x$, $c = 2/3$ for all $\lambda \in [0,1)$ and $\epsilon > 0$. Thus, from lemma 5.3, $A \geq 2Q^*(\epsilon_1)/3$.

Let $x^+(\epsilon_1) \neq x^-(\epsilon_1)$. Then $x^+(\epsilon_1) > x^-(\epsilon_1)$ and $\mathcal{S}_{x^+(\epsilon_1)} < \lambda+\delta(\epsilon_1) < \mathcal{S}_{x^-(\epsilon_1)}$, and $x^+(\epsilon)-x^-(\epsilon) \leq 2\epsilon_1\tilde{r}_{\ell^*}[0]/3$. Note that $(\lambda+\delta(\epsilon_1))\tilde{r}_{\ell^*}[0] \leq Q^*(\epsilon_1)$ because there exists a $\sigma \in \Pi$ which transmits in ℓ^* without probing any channel, and $\mathcal{S}_\sigma = 1$ and $\mathcal{G}_\sigma = (\lambda+\delta(\epsilon_1))\tilde{r}_{\ell^*}[0]$. Thus, $x^+(\epsilon) - x^-(\epsilon) \leq 2/3\frac{\epsilon_1 Q^*(\epsilon_1)}{\lambda+\delta(\epsilon_1)}$. From lemma 5.6, (8) to (10) of lemma 5.4 hold with $\mathcal{L}^+ = x^+(\epsilon_1)$, $\mathcal{L}^- = x^-(\epsilon_1)$, $\hat{\sigma}_{\mathcal{L}^+} = \sigma_{x^+(\epsilon_1)}$, $\hat{\sigma}_{\mathcal{L}^-} = \sigma_{x^-(\epsilon_1)}$, $c = 2/3$ for all $\lambda \in [0,1)$ and $\epsilon > 0$. Thus, from lemma 5.4, $A \geq (2/3)(1-\epsilon_1)Q^*(\epsilon_1) \geq (2/3)(1-\epsilon)Q^*(\epsilon_1)$. The last inequality follows since $\lambda \in [0,1)$. □

References

[1] IEEE 802.11a Working Group. *Wireless LAN Medium Acces Control(MAC) and Physical Layer(PHY) Specifications-Amendment 1:High speed Physical Layer in the 5GHz Band*, 1999.

[2] IEEE 802.11h Working Group. *Wireless LAN Medium Acces Control(MAC) and Physical Layer(PHY) Specifications-Amendment 5:Spectrum and Transmit Power Management Extensions in the 5GHz Band in Europe*, 2003.

[3] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire. The non-stochastic multi-armed bandit problem. *SIAM Journal for Computing*, 32(1):48–77, November 2002.

[4] S. Babu, R. Motwani, K. Munagala, I. Nishizawa, and J. Widom. Adaptive ordering of pipelined stream filters. In *2004*, pages 407–418, June 2004.

[5] D. P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 2nd edition, 2001.

[6] J. Burge, K. Munagala, and U. Srivastava. Ordering pipelined query operators with precedence constraints. *Submitted*, 2005.

[7] P. Chaporkar, K. Kar, and S. Sarkar. Achieving queue length stability through maximal scheduling in wireless networks. In *Proceedings of Information Theory and Applications Inaugural Workshop*, San Diego, CA, April 2006.

[8] Y. S. Chow, H. Robbins, and D. Siegmund. Great expectations: The theory of optimal stopping. *Houghton Mufflin Company*, 1971.

[9] E. Cohen, A. Fiat, and H. Kaplan. Efficient sequences of trials. In *2003*, 2003.

[10] G. Fayolle, V. A. Malyshev, and M. V. Menshikov. *Topics in the Constructive Theory of Countable Markov Chains*. Cambridge University Press, 1995.

[11] U. Feige, L. Lovász, and P. Tetali. Approximating min-sum set cover. *Algorithmica*, 2004.

[12] J. Gittins and D. Jones. A dynamic allocation index for the sequential allocation of experiments. *Progress in Statistics*. edited by J. Gani.

[13] IEEE 802.11 Working Group. *Wireless LAN Medium Acces Control(MAC) and Physical Layer(PHY) Specifications*, 1997.

[14] R. Guha and S. Sarkar. Characterizing temporal snr variation in 802.11 networks. *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, April 3-6 2006.

[15] S. Guha, K. Munagala, and S. Sarkar. Jointly optimal transmission and probing strategies for multichannel wireless systems. In *Proceedings of Conference on Information Sciences and Systems*, Princeton, NJ, March 22-24 2006.

[16] S. Guha, K. Munagala, and S. Sarkar. Optimizing transmission rate in wireless channels using adaptive probes (poster paper). *Proc. of SIGMETRICS/Performance*, 2006.

[17] J. Hellerstein. Optimization techniques for queries with expensive methods. *ACM Transactions on Database Systems*, 23(2):113–157, 1998.

[18] Z. Ji, Y. Yang, J. Zhou, M. Takai, and R. Bagrodia. Exploiting medium access diversity in rate adaptive wireless lans. *ACM MOBICOM*, 2004.

[19] H. Kaplan, E. Kushilevitz, and Y. Mansour. Learning with attribute costs. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 356–365, 2005.

[20] M. S. Kodialam. The throughput of sequential testing. *Lecture Notes in Computer Science*, 2081:280–292, 2001.

[21] K. Munagala, S. Babu, R. Motwani, and J. Widom. The pipelined set cover problem. *Proc. Intl. Conf. Database Theory*, 2005.

[22] K. Munagala, U. Srivastava, and J. Widom. Optimization of continuous queries with shared expensive filters. *Submitted*, 2005.

[23] A. Sabharwal, A. Khoshnevis, and E. Knightly. Opportunistic spectral usage: Bounds and a multi-band csma/ca protocol. *IEEE/ACM Transactions on Networking*, 2006. Accepted for publication.

[24] A. Stolyar, S. Shakkottai, and R. Srikant. Pathwise optimality of the exponential scheduling rule for wireless channels. *Advances in Applied Probability*, 36(4):1021–1045, 2004.

[25] E. Vergetis, R. Guerin, and S. Sarkar. Realizing the benefits of user-level channel diversity. *Computer Communications Review*, 35(5), October 2005.